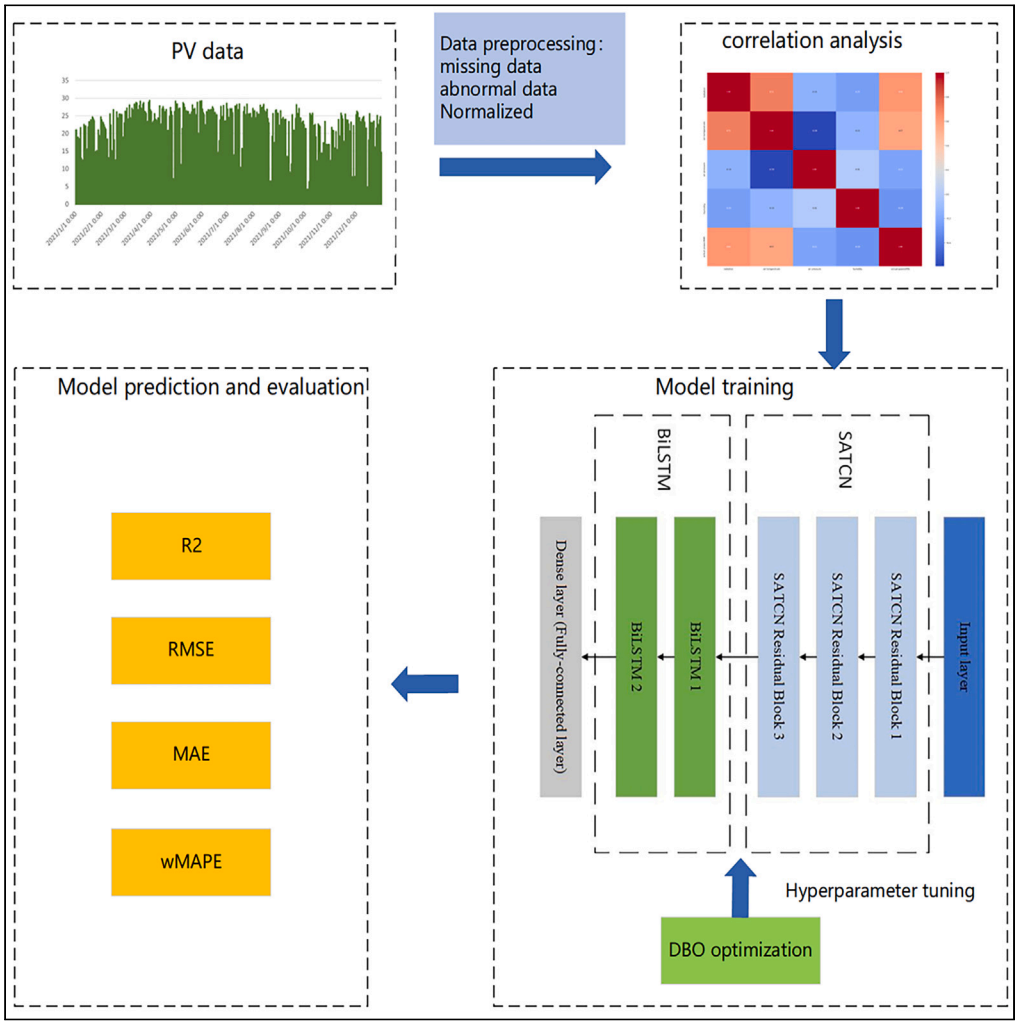
iScience

CellPress
OPEN ACCESS

Article

Dung beetle optimization algorithm-based hybrid deep learning model for ultra-short-term PV power prediction



Rui Quan, Zhizhuo Qiu, Hang Wan, Zhiyu Yang, Xuerong Li

102210382@hbut.edu.cn

Highlights

The proposed SATCN-BiLSTM model presents a more precise approach for PV power prediction

Unique DBO optimization algorithm optimizes model hyperparameters

The proposed model can be used for ultra-short-term PV power prediction

Compare and evaluate the proposed model with existing models and optimization algorithms

Quan et al., iScience 27, 111126
November 15, 2024
https://doi.org/10.1016/j.isci.2024.111126

iScience

CellPress
OPEN ACCESS

Article

Dung beetle optimization algorithm-based hybrid deep learning model for ultra-short-term PV power prediction

Rui Quan,[1,2] Zhizhuo Qiu,[1,2,3,*] Hang Wan,[1,2] Zhiyu Yang,[1,2] and Xuerong Li[1,2]

SUMMARY

A hybrid model combining self-attention temporal convolutional networks (SATCN) with bidirectional long short-term memory (BiLSTM) networks was developed to improve the accuracy of ultra-short-term photovoltaic (PV) power prediction. The self-attention mechanism and SATCN were used to extract temporal and correlation features, which were then linked to BiLSTM networks. The model's hyperparameters were optimized using the dung beetle optimization algorithm. The model was tested on a year-long dataset of PV power and outperformed convolutional neural networks, BiLSTM networks, temporal convolutional networks, and other hybrid models. It reduced the root-mean-square error (RMSE) by 33.1% compared to the other models. The model achieved a mean absolute error (MAE) of 0.175, a weighted mean absolute percentage error (wMAPE) of 4.821, and a coefficient of determination (R2) of 0.997. These results highlight the model's superior accuracy and its potential applications in solar energy development.

INTRODUCTION

The surging global energy demand and the escalating climate change crisis have propelled renewable energy to the forefront of the energy development landscape. Among these renewable sources, photovoltaic (PV) power generation is a crucial element.[1] The gradual integration of PV power systems into the power grid has become customary.[2] The reliance of PV installations on solar radiation introduces distinctive challenges due to their intermittent nature and susceptibility to weather conditions, hindering their reliable and efficient operation.[3] The intermittent and fluctuating characteristics of PV power generation can significantly impact the main grid, posing substantial risks to grid functionality.[4] Accurate PV power forecasting enables real-time decision-making by operators, reduces grid security risks, and enhances the quality and economic efficiency of power supply.[5]

Regarding power forecasting techniques, PV power prediction methods are typically classified into distinct groups: physical, statistical, meta-heuristic learning, and hybrid methodologies.[6] The physical approach involves building a realistic model of solar radiation probability distribution through theoretical modeling, such as Normal, Weibull, and Beta distributions.[7] This method selects appropriate parameters and constructs a probabilistic forecast model for solar radiation intensity, leveraging the inherent link between solar radiation and PV power prediction.[8] Physical methods require less long-term information and are well-suited for long-term forecasts. They rely on accurate numerical models and geographic-specific information, making their establishment and application challenging.[9] In contrast to physical methods, statistical methods offer the advantages of simpler modeling and broader applicability. These methods establish mapping connections between historical data on solar radiation, wind speed, and PV power through curve fitting, parameter estimation, and correlation analysis to achieve accurate predictions.[10] Statistical methods encompass time series analysis, regression analysis, fuzzy theory, and gray theory.[11] The effectiveness of these methods hinges on abundant, meticulously processed historical data, and their prediction accuracy is intricately linked to data quality, data volume, screening, and error removal.[12]

Meta-heuristic learning methods typically involve applying neural networks and relevant algorithms to establish relationships between inputs and outputs for prediction, becoming increasingly prevalent in PV power prediction.[13] Many scholars have achieved a large number of results using classical convolutional neural networks (CNNs),[14] recurrent neural networks (RNN),[15] and long short-term memory (LSTM) networks.[16] Bai et al.[17] proposed a temporal convolutional neural network (TCN) based on CNNs, showing that convolutional architectures outperform the typical recurrent networks on a wide range of tasks and datasets, with a flexible sensory field demonstrating a longer effective memory. Quan et al.[18] proposed deep bidirectional LSTM (BiLSTM) network modeling, applying LSTM networks twice to improve long-term dependency learning and enhance model accuracy.[19] Numerous research findings highlight that the predictive accuracy of a combined model, integrating multiple algorithms or models, notably surpasses that of a singular prediction model. For instance, Zou et al.[20] employed

[1]Hubei Key Laboratory for High-efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan 430068, China
[2]Hubei Engineering Research Center for Safety Monitoring of New Energy and Power Grid Equipment, Hubei University of Technology, Wuhan 430068, China
[3]Lead contact
*Correspondence: 102210382@hbut.edu.cn
https://doi.org/10.1016/j.isci.2024.111126

a CNN and BiLSTM network fusion in wind power prediction, effectively leveraging temporal data features to enhance predictive accuracy. Limouni et al.[21] introduced an LSTM-TCN model for PV power prediction, which utilizes LSTM to analyze the time domain characteristics of the input signal and integrate it with TCN, demonstrating superior predictive performance compared to single-model approaches. While current research on combined prediction models emphasizes using deep learning for data feature extraction and power prediction, it's important to note that these network parameters are determined empirically, and tuning them does not guarantee optimal model performance in every scenario.

The fusion of swarm intelligence algorithms for hyperparameter tuning of PV power prediction models has emerged as a popular and extensively explored field. These algorithms encompass genetic algorithm (GA), particle swarm optimization (PSO), dung beetle optimization (DBO) algorithm, sparrow search algorithm (SSA),[22] and so on. The cat swarm optimization (CSO) algorithm was utilized by Liu to identify optimal parameters for boosting the forecast precision of the backpropagation (BP) neural network.[23] Wu et al.[24] utilized the ant colony optimization (ACO) to optimize support vector machine (SVM) parameters, which enhanced the forecasting precision of the BP neural network through a series of improvements, resulting in a 6.8% enhancement in the model's regression coefficient of determination (R2) via judicious data preprocessing. Abou et al.[25] build a COA-CNN-LSTM model, integrating the Coati optimization algorithm to enhance hyperparameters, consequently improving the model's rate of learning. These efforts underscore the potential of swarm intelligence algorithms to improve neural network models and enhance predictive accuracy in PV power forecasting tasks.

However, existing PV power generation prediction methods have several shortcomings. Physical methods are heavily dependent on accurate models and specific geographic data, statistical methods require a large amount of historical data. Meta-heuristic learning methods (such as neural networks) are complex and time-consuming to train, and often rely on empirical parameter adjustment, which cannot obtain the optimal results. Swarm intelligence algorithms are expected to improve the optimal parameters of neural network models, but the optimization algorithms currently used for PV power prediction still have certain limitations. For example, the optimization methods of optimization algorithms such as PSO and SSA are relatively simple and easy to fall into local optimality. To address the aforementioned issues, this paper proposes a DBO-SATCN-BiLSTM model to solve the problem of ultra-short-term PV power prediction. Self-attention temporal convolutional network (SATCN) combines the self-attention mechanism and the temporal convolutional network; self-attention is used to capture the dependencies between different time steps in time series data, while temporal convolution is used to learn local and global temporal features. The BiLSTM layer uses its bidirectional data processing capabilities to obtain the forward and backward data change characteristics in the sequence data, and the hyperparameters of the SATCN-BiLSTM model are optimized using the DBO algorithm. The proposed model is then compared with single models such as CNN, BiLSTM and TCN, hybrid models such as CNN-BiLSTM and TCN-BiLSTM, and models with different optimization algorithms such as PSO-SATCN-BiLSTM and SSA-SATCN-BiLSTM. The purpose of this comparative analysis is to verify the advantages and accuracy of the model. The main contributions of this work include the following.

(1) A hybrid model combing SATCN with BiLSTM was proposed to predict ultra-short-term PV power.
(2) The distinctive DBO optimization algorithm collaborates with SATCN-BiLSTM (DBO-SATCN-BiLSTM) to improve and the model structure and optimize its hyperparameters through refinement and fine-tuning.
(3) The DBO-SATCN-BiLSTM model is effectively applied for ultra-short-term single-step as well as multi-step prediction of PV power.
(4) The proposed model performs better than existing traditional single and hybrid models optimized with various other optimization algorithms.

The following sections are organized as follows. Section "results and discussion" presents and analyzes the results of ultra-short-term single-step and multi-step PV power prediction experiments in detail. Section "system structure and model" introduces the SATCN, BiLSTM, and SATCN-BiLSTM model structures and the DBO optimization algorithm. The proposed DBO-SATCN-BiLSTM workflow is introduced. Section "data preprocessing and model evaluation" introduces the data preprocessing methods and the evaluation metrics of the models. Section "conclusion" presents the conclusions of the study. The article ends with detailed methods.

RESULTS AND DISCUSSION

Parameter settings

Initially, the hyperparameters of SATCN-BiLSTM are determined using the DBO algorithm. SATCN-BiLSTM relies on eight critical parameters: neurons1, neurons2, neurons3, kernel_size1, kernel_size2, kernel_size3, cell_size1, and cell_size2. To validate the optimization algorithms' feasibility, three comparison experiments—DBO-SATCN-BiLSTM, SSA-SATCN-BiLSTM, and PSO-SATCN-BiLSTM—are conducted separately. The specific parameters of the optimization algorithm are outlined in Table 1. The iteration curves are shown in Figure 1, the DBO algorithm reaches a stable state at 11 iterations and has the smallest average error, the PSO tends to be stable at 15 iterations and has the largest single average error, and the SSA reaches a stable state at 23 iterations, with an average error lower than the PSO but higher than that of the DBO. The DBO algorithm has the smallest average error compared with the PSO and SSA algorithms and the convergence speed is faster.

To verify the efficiency of the DBO-SATCN-BiLSTM model, it is compared with BiLSTM, TCN, CNN, TCN-BiLSTM, CNN-BiLSTM, PSO-SATCN-BiLSTM, and SSA-SATCN-BiLSTM using the aforementioned evaluation metrics. The detailed setup training model parameters are shown in Table 2.

Table 1. Optimize algorithm parameter settings

Algorithm	Iteration	Population size	Proportional parameter
PSO	50	10	Inertia weight:0.5; Learning factors:1.5, 1.6
SSA	50	10	Warning value: Discoverer: conscious = 0.6: 0.7: 0.3
DBO	50	10	Ball: brood: baby: thief = 0.4: 0.6: 0.4: 0.4

Annual prediction performance evaluation

Throughout the annual dataset, eight models undergo one-step prediction individually, and the prediction curves for these models are depicted in Figures 2 and 3. Analysis reveals that all eight models exhibit errors in predicting PV power, yet the DBO-SATCN-BiLSTM model with smaller errors and reduced fluctuation ranges. Notably, the predicted values using DBO-SATCN-BiLSTM (marked in green) closely align with the actual power (depicted in black). Even amid intermittent PV power generation scenarios, this model maintains a high level of stability.

Table 3 presents the results of one-year PV power prediction using the DBO-SATCN-BiLSTM model. The root-mean-square error (RMSE) stands at 0.357, showcasing a 33.1%, 23.4%, 18.1%, 17.5%, and 10% reduction compared to CNN, BiLSTM, TCN, CNN-BiLSTM, and TCN-BiLSTM. It demonstrates 2.9% and 2.4% decrease in comparison to PSO-SATCN-BiLSTM and SSA-SATCN-BiLSTM. The mean absolute error (MAE) value of 0.175 marks a reduction of 36.1%, 26.4%, 26.7%, 23.5%, and 13.7% in comparison to the five aforementioned models , and decreases by 9.7% and 8.3% compared to PSO-SATCN-BiLSTM and SSA-SATCN-BiLSTM. The trend is visualized in Figure 4, highlighting the superior optimization effect of DBO over PSO and SSA in the SATCN-BiLSTM model. The DBO algorithm excels due to its rapid convergence and high accuracy, leveraging both global exploration and local exploitation. The smallest weighted mean absolute percentage error (wMAPE) value records at 4.821%, while the R2 value of 0.997 indicates exceptional fitting accuracy.

In multi-step prediction, all models experienced decreased prediction accuracy, evident in the prediction curves displayed in Figures 5 and 6. Among them, the three single models—CNN, BiLSTM, and TCN—exhibit the most significant increase in errors. Following these are the hybrid models CNN-BiLSTM and TCN-BiLSTM. The SATCN-BiLSTM model optimized with algorithms showed decreased errors while maintaining good stability compared to CNN, BiLSTM, TCN, CNN-BiLSTM, and TCN-BiLSTM. For the DBO-SATCN-BiLSTM model, the RMSE, MAE, and wMAPE are 0.437, 0.217, and 5.966, respectively, marking these values as the smallest among all the models considered. The R2 value is the largest compared to the other models. This validates that the prediction accuracy of the DBO-SATCN-BiLSTM model surpasses that of the other seven models in multi-step prediction scenarios.

Comparative results of single-step forecasts in different seasons

For the single-step PV power prediction, it includes only one prediction in the output window, and Figure 7 illustrates the single-step prediction curves for PV power across spring, summer, fall, and winter. An observed trend among all models is lower errors in winter and fall, contrasted with higher prediction errors in spring and summer. This trend aligns with the local weather patterns: the winter season typically experiences less precipitation and more sunny days, resulting in relatively stable PV curves. In contrast, spring and summer, characterized by rainy seasons, display more weather variability, and contribute to increased errors in predicting PV power.

Table 4 presents the RMSE, MAE, wMAPE, and R2 values for single-step predictions from three single models (CNN, BiLSTM, and TCN), hybrid models (CNN-BiLSTM and TCN-BiLSTM), and three different optimized SATCN-BiLSTM models (DBO-SATCN-BiLSTM,

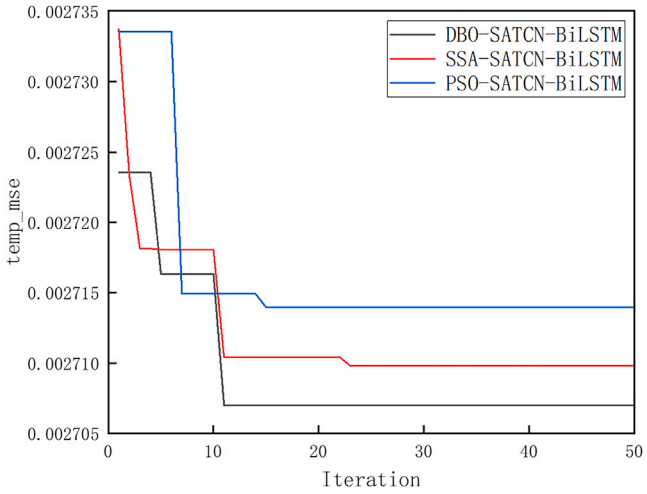


Figure 1. Fitness of different optimization algorithms to the model

Table 2. Model training parameter

Model	Parameters
CNN	Filters1 = 128; Filters2 = 64; Filters3 = 32; Kernels1 = 3; Kernels2 = 3; Kernels3 = 3;
BiLSTM	cell_size1 = 128; cell_size2 = 64;
TCN	Filters1 = 128; Filters2 = 64; Filters3 = 32; Kernels1 = 3; Kernels2 = 3; Kernels3 = 3; Dilation1 = 6; Dilation2 = 3; Dilation3 = 2;
CNN-BiLSTM	Filters1 = 128; Filters2 = 64; Filters3 = 32; Kernels1 = 3; Kernels2 = 3; Kernels3 = 3; cell_size1 = 128; cell_size2 = 64;
TCN-BiLSTM	Filters1 = 128; Filters2 = 64; Filters3 = 32; Kernels1 = 3; Kernels2 = 3; Kernels3 = 3; Dilation1 = 6; Dilation2 = 3; Dilation3 = 2; cell_size1 = 128; cell_size2 = 64;
PSO-SATCN-BiLSTM	Filters1 = 86; Filters2 = 19; Filters3 = 58; Kernels1 = 33; Kernels2 = 29; Kernels3 = 15; Dilation1 = 6; Dilation2 = 3; Dilation3 = 2; cell_size1 = 114; cell_size2 = 195;
SSA-SATCN-BiLSTM	Filters1 = 23; Filters2 = 79; Filters3 = 103; Kernels1 = 41; Kernels2 = 17; Kernels3 = 67; Dilation1 = 6; Dilation2 = 3; Dilation3 = 2; cell_size1 = 191; cell_size2 = 103;
DBO-SATCN-BiLSTM	Filters1 = 83; Filters2 = 119; Filters3 = 96; Kernels1 = 36; Kernels2 = 12; Kernels3 = 34; Dilation1 = 6; Dilation2 = 3; Dilation3 = 2; cell_size1 = 23; cell_size2 = 199;
Common parameters	
Epoch	50
Batch_size	64
Dropout	0.3
Optimizer	Adam
Learning rate	0.001
Loss	MSE

SSA-SATCN-BiLSTM, and PSO-SATCN-BiLSTM) across each season within a year. During the winter season, DBO-SATCN-BiLSTM demonstrates lower prediction errors with an RMSE of 0.329. Compared to CNN, BiLSTM, TCN, CNN-BiLSTM, TCN-BiLSTM, PSO-SATCN-BiLSTM, and SSA-SATCN-BiLSTM, it is decreased by 35.8%, 25.9%, 32%, 14.3%, 17.1%, 14.1%, and 9.8%. DBO-SATCN-BiLSTM exhibits a small MAE of 0.183, which is reduced by 43.6%, 21.4%, 28.7%, 19.3%, 18.6%, 14.4%, and 11.5%, compared to the aforementioned models. This underscores DBO's superior optimization effect on SATCN-BiLSTM. It showcases lower wMAPE of 4.575 and a higher R2 value of 0.998, indicating better predictive accuracy. During the summer season, where errors tend to be higher across models, DBO-SATCN-BiLSTM maintains favorable performance, showcasing an RMSE of 0.723, an MAE of 0.355, a wMAPE of 7.16, and an R2 of 0.991. These values outperform those of the other seven models, emphasizing its robustness in predicting PV power during the challenging summer season.

The strength of the DBO-SATCN-BiLSTM model lies in its capacity to extract features not only from adjacent days but also from historical days exhibiting similar weather patterns, which effectively enhances the model's ability to generalize. In comparison to other models, the proposed model consistently demonstrates lower prediction errors and better fits to real curves. This robust performance across various seasons signifies the resilience and adaptability of the DBO-SATCN-BiLSTM model to diverse environmental conditions.

Comparative results of multi-step forecasts in different seasons

For the multi-step PV power prediction, the output window contains multiple consecutive values, and Figure 8 plots the multi-step (3-step) prediction curves for PV power across the same four seasons. Observations indicate an increase in deviation between prediction curves and actual curves, and multi-step predictions of CNN, BiLSTM, and TCN show higher deviations. This trend continues with hybrid models of CNN-BiLSTM and TCN-BiLSTM, and gradually increases in deviation from the actual curves. SATCN-BiLSTM demonstrates a comparatively small increase in deviation from the actual curve, maintaining relative stability.

Table 5 lists the RMSE, MAE, wMAPE and R2 values for multi-step predictions across each model for the aforementioned four seasons. As anticipated, the multi-step error values incrementally raised compared to single-step predictions. In this case, the smallest prediction errors persist during winter, with DBO-SATCN-BiLSTM exhibiting an RMSE value of 0.416, which marks a 35.4%, 33.5% and 33.2% reduction compared to CNN, TCN, and BiLSTM. DBO-SATCN-BiLSTM shows a 33.9% decrease compared to hybrid models of CNN-BiLSTM and TCN-BiLSTM, and 29.7%, 14.0%, and 12.4% reductions compared to PSO-SATCN-BiLSTM and SSA-SATCN-BiLSTM. The MAE value of DBO-SATCN-BiLSTM (0.286) is notably decreased by 17.5%, 13.8%, 16.1%, 11.2%, 13.3%, 9.4%, and 6.8%—than CNN, BiLSTM, TCN, CNN-BiLSTM, TCN-BiLSTM, PSO-SATCN-BiLSTM, and SSA-SATCN-BiLSTM. Similarly, the wMAPE value of DBO-SATCN-BiLSTM is also lower than other models, showcasing better predictive accuracy, while the R2 value of 0.998 signifies a strong fit. Despite the substantial errors

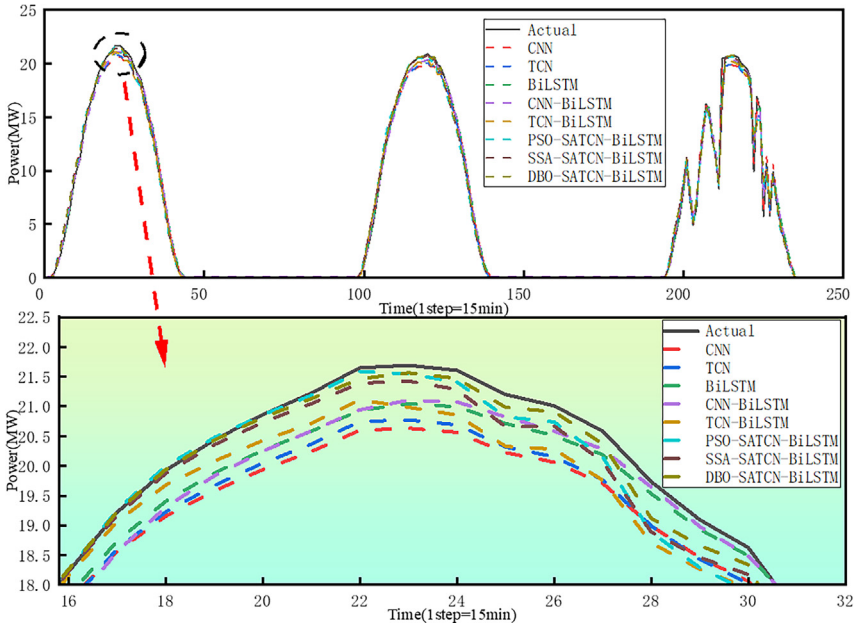


Figure 2. Smoothed single-step PV power prediction curve for any day of a year

observed in summer, the SATCN-BiLSTM model optimized with optimization algorithms maintains a high prediction accuracy compared to single models of CNN, TCN, and BiLSTM, showing resilience to weather fluctuations. In contrast, the prediction accuracies of the other five models without optimization fluctuate significantly, which further underscores the superior optimization effect of DBO over PSO and SSA algorithms.

Single model's prediction accuracy significantly lags behind a hybrid model's when handling fluctuating prediction samples. The challenge involves the single prediction model's limited capacity to effectively discern fluctuation patterns within historical data, resulting in diminished prediction accuracy. Conversely, hybrid models, especially those fine-tuned with intelligent optimization algorithms, consistently demonstrate superior prediction accuracy compared to their non-optimized counterparts.

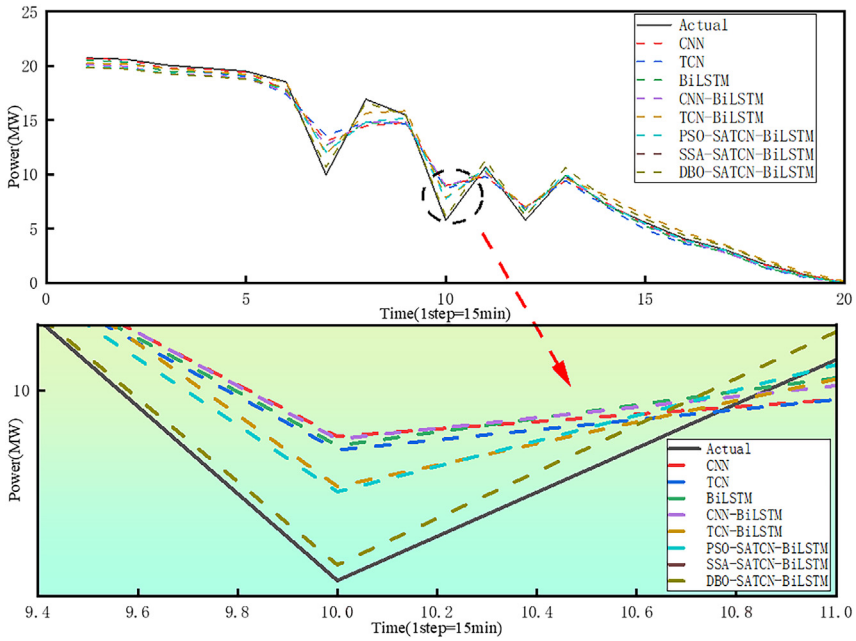


Figure 3. Abrupt single-step PV power prediction curve for any day of a year

Table 3. Single-step and multi-step PV power prediction errors for the annual

Step	Model	RMSE	MAE	wMAPE	R2
1 (15 min)	CNN	0.534	0.274	7.57	0.993
	BiLSTM	0.466	0.238	6.568	0.995
	TCN	0.436	0.239	6.311	0.996
	CNN-BiLSTM	0.433	0.229	5.766	0.995
	TCN-BiLSTM	0.4	0.203	5.606	0.996
	PSO-SATCN-BiLSTM	0.368	0.194	5.102	0.997
	SSA-SATCN-BiLSTM	0.366	0.191	4.931	0.997
	DBO-SATCN-BiLSTM	0.357	0.175	4.821	0.997
3 (45 min)	CNN	0.917	0.661	10.406	0.981
	BiLSTM	0.647	0.376	8.919	0.99
	TCN	0.652	0.377	8.894	0.99
	CNN-BiLSTM	0.638	0.323	8.329	0.991
	TCN-BiLSTM	0.635	0.302	8.137	0.99
	PSO-SATCN-BiLSTM	0.483	0.274	6.167	0.995
	SSA-SATCN-BiLSTM	0.459	0.254	6.003	0.995
	DBO-SATCN-BiLSTM	0.437	0.217	5.966	0.995

Public dataset testing

To verify the applicability of the proposed model with different datasets, a set of public datasets was included in the experiment. The public dataset originates from the China State Grid Renewable Energy Generation Forecasting Competition Project,[26] which comprises PV power generation-related data from 2019 to 2020, with a data collection interval of 15 min. The collected features include solar radiation, atmospheric temperature, cloud thickness, and other relevant information. The PV power plant from which the data were collected is located in northern China and has an installed capacity of 50 MW Table 6 lists the relevant information about the dataset, including the mean, maximum, minimum, and standard deviation of the feature data.

In the dataset spanning two years, eight models were used for single-step prediction experiments (15 min) and multi-step prediction experiments (45 min). Figure 9 shows the prediction curves of the DBO-SATCN-BiLSTM model and other comparison models in single-step prediction. Figure 9A extracts 175 sample points from the test set for display, including cloudy and rainy days (24–75 steps) and sunny days (125–175 steps). Figure 9B is a local enlargement of the sample points from 130 to 155. It can be seen that the DBO-SATCN-BiLSTM model, represented by the green curve, is always closer to the true value curve than other comparison models. When the SATCN-BiLSTM model remains unchanged, the overall prediction accuracy of the model optimized by the DBO optimization algorithm is higher than that of the model optimized by PSO or SSA. The same trend is observed in multi-step prediction. Figures 10A and 10B show the prediction curves of each model in multi-step prediction. Due to the larger time span of multi-step prediction, the prediction curve of each model is slightly less

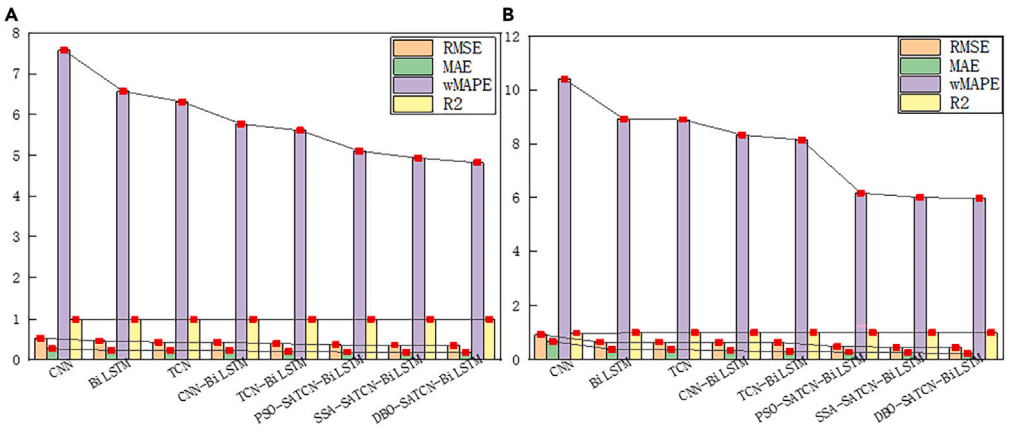


Figure 4. Power prediction error of annual data

(A) Single-step.

(B) Multi-step.

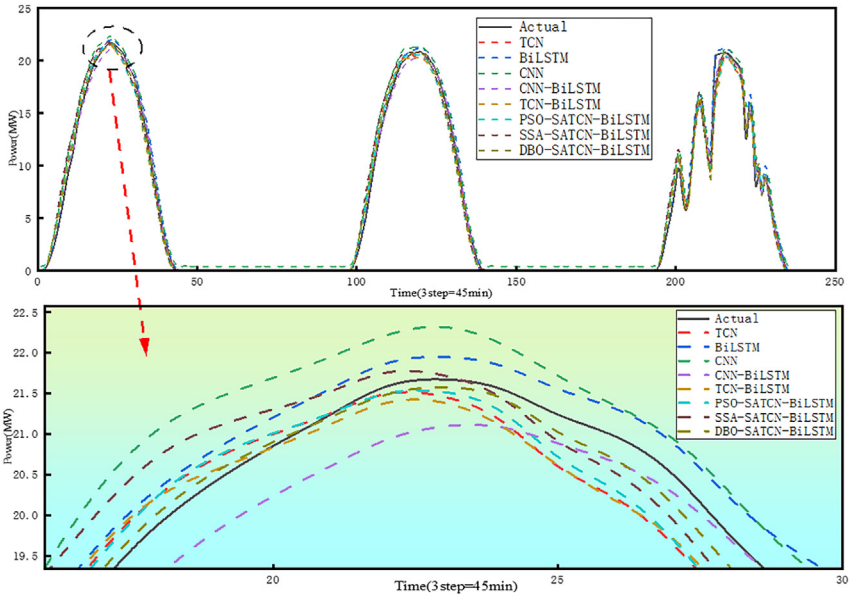


Figure 5. Smoothed multi-step PV power prediction curve for any day of a year

consistent with the true value curve than in single-step prediction. Nonetheless, the prediction curve of DBO-SATCN-BiLSTM remains closely aligned with the true value overall, demonstrating more accurate prediction performance than other models.

Table 7 shows the evaluation indicators of all models in single-step and multi-step predictions. The evaluation indicators include RMSE, MAE, wMAPE, and R2. For single-step prediction, the RMSE, MAE, wMAPE, and R2 of the proposed model are 0.576, 0.438, 5.294, and 0.994. These errors are significantly smaller than those of other models such as CNN, BiLSTM, TCN, CNN-BiLSTM, and TCN-BiLSTM. The fitting value R2 of the proposed model is higher than that of the single models. Specifically, compared with the CNN model, the proposed model's errors were reduced by 43.91%, 56.29%, and 43.95%, and R2 was increased by 0.006. Compared with the TCN-BiLSTM model, its errors were reduced by 28.09%, 37.69%, and 27.82%, and R2 was increased by 0.002.

The proposed model is improved by the self-attention mechanism. TCN can focus more on the correlation between each data point in the sequence and all other data points, thereby dynamically assigning weights to each position to integrate the sequence information. General TCN usually defines a fixed receptive field through the size and number of layers of the convolution kernel, which somewhat limits their ability to capture long-term dependencies.[27] Compared with optimization algorithms such as PSO and SSA, the DBO optimization algorithm achieves lower prediction errors and higher fitting values. In terms of wMAPE, the DBO algorithm is 21.36% lower than

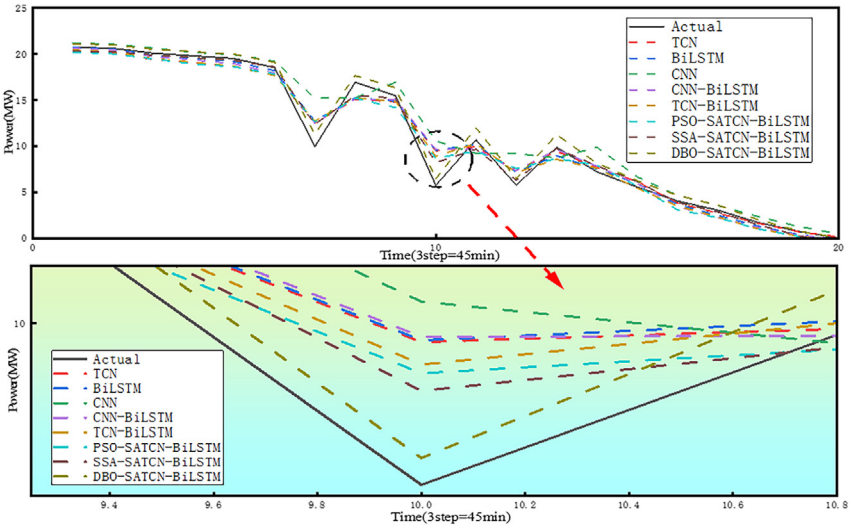


Figure 6. Abrupt multi-step PV power prediction curve for any day of a year

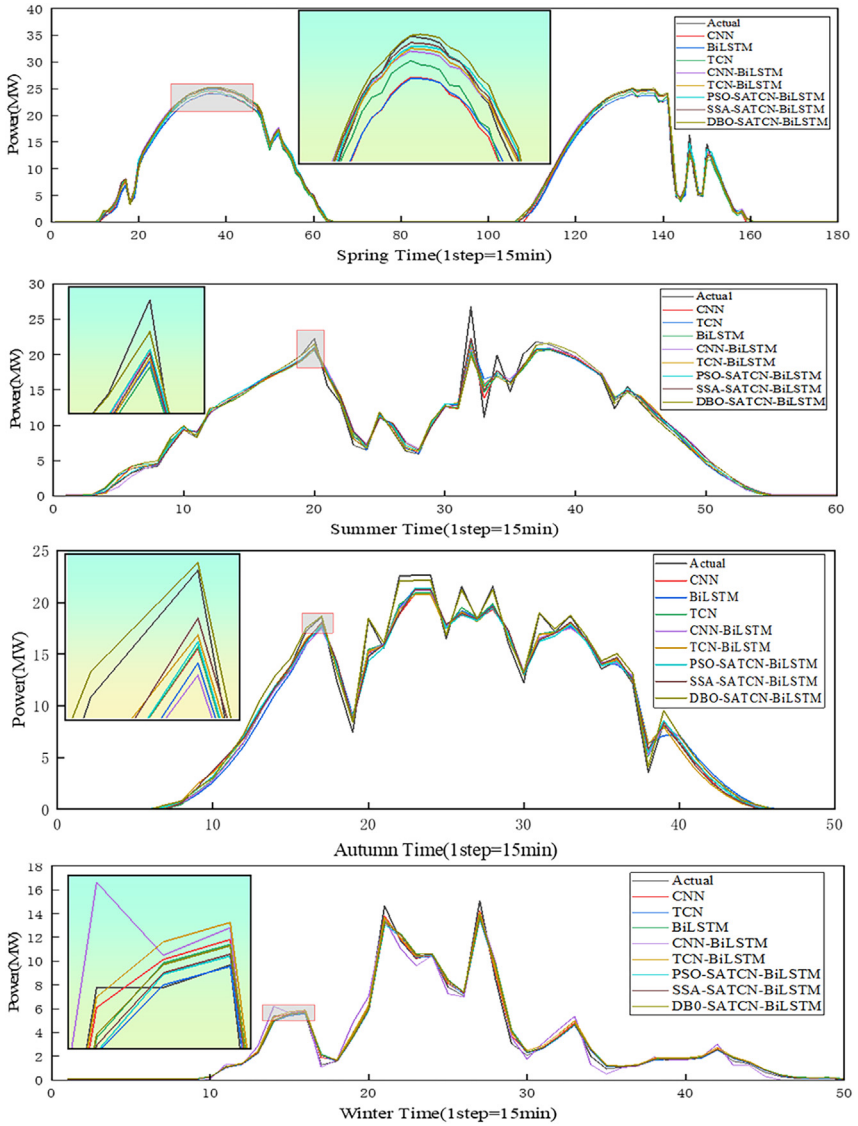


Figure 7. Single-step (step1 = 15 min) PV power prediction curves for four seasons

PSO and 20.93% lower than SSA. These lower errors and higher fitting values demonstrate the superiority of the proposed model in prediction.

In multi-step prediction, the error of each model increases significantly, and the fitting values decrease. The DBO-SATCN-BiLSTM model still maintains high stability and accuracy. Specifically, in the 45-min multi-step prediction, the RMSE, MAE, wMAPE, and R^2 of the DBO-SATCN-BiLSTM model are 0.722, 0.613, 6.849, and 0.991. These errors are significantly lower than those of models without an optimization algorithm. The proposed model's error is smaller than that of the other two optimization algorithms compared against it. Compared with PSO-SATCN-BiLSTM, RMSE, MAE, and wMAPE are reduced by 10.97%, 22.60%, and 6.14%. Compared with SSA-SATCN-BiLSTM, these metrics are reduced by 10.31%, 20.90%, and 6.74%. The use of the DBO optimization algorithm has significantly improved the model's prediction accuracy. This improvement is attributed to the DBO algorithm's integration of various optimization strategies, which enable the algorithm to explore the solution space in different ways simultaneously. DBO introduces a pheromone model similar to the ant colony algorithm. This model helps the dung beetle remember successful search paths and adjust the search strategy by updating the pheromone concentration, thereby increasing the algorithm's global search capabilities and convergence speed. Although SSA and PSO can also perform global searches, their search strategies and update mechanisms may be relatively simple, relying on the position and speed of particles or sparrows for updates,[28] which may cause them to fall into local optimal solutions in some cases.

Table 4. Single-step PV power prediction errors for four seasons with different models

Seasons	Model	RMSE	MAE	wMAPE	R2
Spring	CNN	0.837	0.47	6.744	0.992
	BiLSTM	0.731	0.433	6.267	0.993
	TCN	0.787	0.359	5.629	0.992
	CNN-BiLSTM	0.794	0.358	5.18	0.993
	TCN-BiLSTM	0.713	0.402	5.812	0.994
	PSO-SATCN-BiLSTM	0.654	0.389	5.2	0.994
	SSA-SATCN-BiLSTM	0.639	0.354	5.129	0.994
	DBO-SATCN-BiLSTM	**0.569**	**0.315**	**4.558**	**0.996**
Summer	CNN	0.862	0.455	9.176	0.987
	BiLSTM	0.867	0.417	8.415	0.987
	TCN	0.837	0.365	7.371	0.988
	CNN-BiLSTM	0.902	0.535	10.79	0.986
	TCN-BiLSTM	0.844	0.417	8.41	0.988
	PSO-SATCN-BiLSTM	0.799	0.371	7.485	0.989
	SSA-SATCN-BiLSTM	0.75	0.369	7.443	0.99
	DBO-SATCN-BiLSTM	**0.723**	**0.355**	**7.16**	**0.991**
Autumn	CNN	0.613	0.255	7.472	0.992
	BiLSTM	0.58	0.243	7.117	0.993
	TCN	0.624	0.266	7.79	0.991
	CNN-BiLSTM	0.538	0.25	7.306	0.994
	TCN-BiLSTM	0.53	0.239	7.008	0.994
	PSO-SATCN-BiLSTM	0.505	0.242	7.079	0.994
	SSA-SATCN-BiLSTM	0.481	0.237	6.925	0.995
	DBO-SATCN-BiLSTM	**0.339**	**0.18**	**5.263**	**0.997**
Winter	CNN	0.513	0.325	8.144	0.995
	BiLSTM	0.444	0.233	6.072	0.996
	TCN	0.484	0.257	6.425	0.996
	CNN-BiLSTM	0.384	0.227	5.677	0.997
	TCN-BiLSTM	0.397	0.225	5.687	0.997
	PSO-SATCN-BiLSTM	0.386	0.214	5.621	0.997
	SSA-SATCN-BiLSTM	0.365	0.207	5.175	0.997
	DBO-SATCN-BiLSTM	**0.329**	**0.183**	**4.575**	**0.998**

The low error and high fit of the proposed model in both single-step and multi-step predictions verify its applicability in different application scenarios. The experiments in the public measured dataset also provide data reference for the practical application of the proposed model.

System structure and model

Self-attention temporal convolutional network

As depicted in Figure 11, the proposed SATCN's residual structure incorporates TCN—a CNN customized for the analysis of time series data. TCN employs dilation convolution kernels and a residual structure to capture global sequence information while mitigating issues like gradient vanishing or explosion.[29] Within the residual structure module, the input undergoes dilation convolution, weight normalization, activation via ReLU and two rounds of dropout operations. Dropout enhances training speed and guards against overfitting. The residual link encompasses two convolutional units and nonlinear mapping. Expansion convolution is employed in the convolutional segment, adjusting the sampling interval via one-dimensional expansion coefficients to enable the network to retain sufficient historical information. In the enhanced SATCN residual structure, a self-attention mechanism is introduced post each dilation convolution. This mechanism autonomously learns and computes the input-output contribution, enabling the model to emphasize crucial segments. This augmentation significantly bolsters the model's capacity to capture feature information from pivotal points within the PV power sequence.

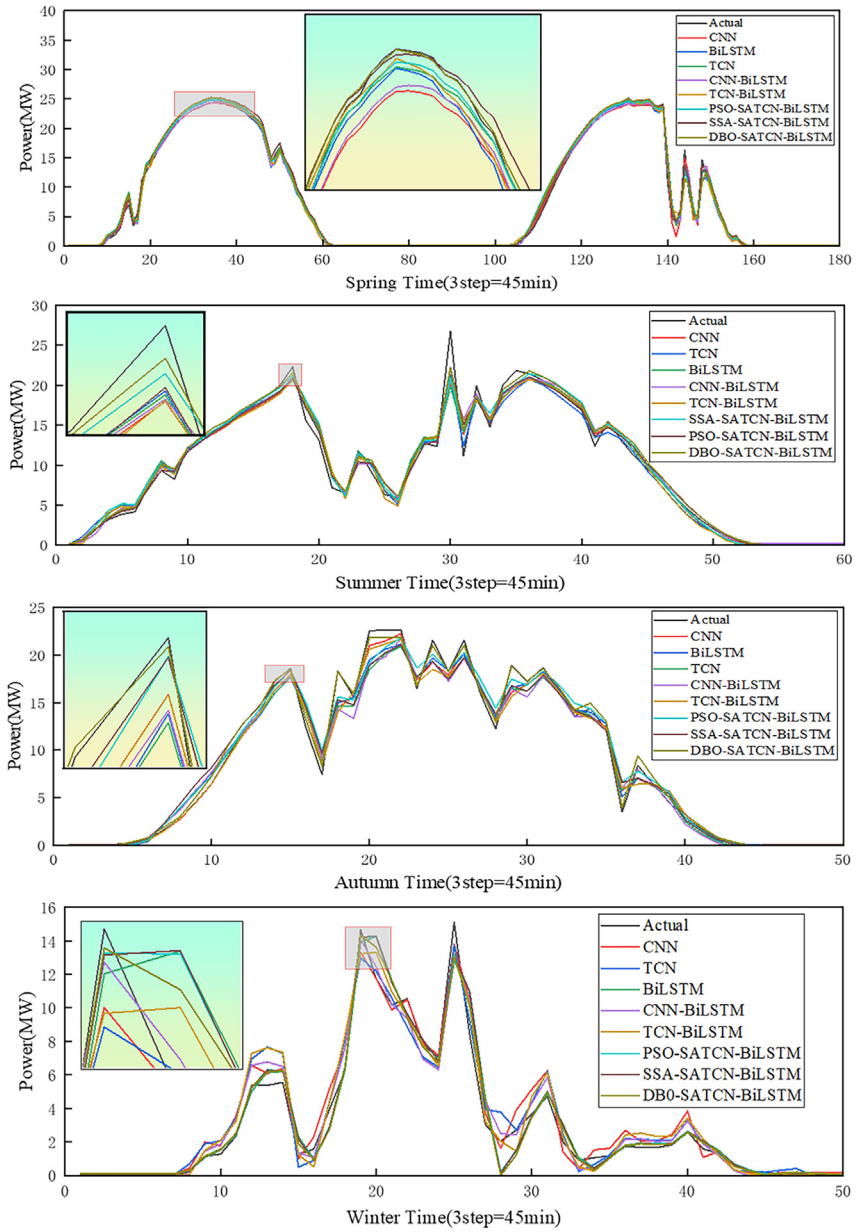


Figure 8. Multi-step (step3 = 45 min) PV power prediction curves for four seasons

To tackle the constrained perceptual domain in causal convolution, an innovative solution named extended convolution is introduced. Figure 12 illustrates the framework of this extended convolution network. Here, the input $x_0, x_1,\ldots x_{t-1}$ from a historical moment is denoted as x_t, serving as the model's input at the moment t, meeting the prerequisites for time series prediction. The resulting output is labeled as $y_0, y_1,\ldots y_t$, while d signifies the number of expansions integrated into the system. The extended convolution can handle longer temporal data dependencies at the same complexity, and increase the model's predictive accuracy.

The output x dilation convolution post dilation factor is given in the following equation:

$$F(s) = \sum_{i=0}^{k-1} f(i) x_{s-di} \qquad \text{(Equation 1)}$$

where $F(s)$ represents the dilation convolution, k denotes the convolution kernel size, $f(i)$ signifies the ith element within the convolution kernel, and $x_{s\text{-}di}$ represents the sequence element that corresponds to the multiplication with the convolution kernel.

Table 5. Multi-step PV power prediction errors for four seasons with different models

Seasons	Model	RMSE	MAE	wMAPE	R2
Spring	CNN	0.981	0.496	7.205	0.988
	BiLSTM	0.912	0.493	7.162	0.989
	TCN	0.828	0.458	6.651	0.991
	CNN-BiLSTM	0.828	0.441	6.395	0.991
	TCN-BiLSTM	0.835	0.466	6.76	0.991
	PSO-SATCN-BiLSTM	0.722	0.421	5.913	0.994
	SSA-SATCN-BiLSTM	0.714	0.407	5.938	0.993
	DBO-SATCN-BiLSTM	**0.694**	**0.409**	**5.896**	**0.994**
Summer	CNN	1.165	0.535	10.857	0.984
	BiLSTM	0.947	0.468	9.493	0.984
	TCN	0.951	0.433	8.772	0.984
	CNN-BiLSTM	0.877	0.436	8.649	0.987
	TCN-BiLSTM	0.894	0.437	8.671	0.986
	PSO-SATCN-BiLSTM	0.833	0.406	8.233	0.988
	SSA-SATCN-BiLSTM	0.783	0.367	7.436	0.989
	DBO-SATCN-BiLSTM	**0.765**	**0.392**	**7.946**	**0.99**
Autumn	CNN	0.667	0.363	10.607	0.99
	BiLSTM	0.595	0.239	7.002	0.992
	TCN	0.622	0.277	8.115	0.991
	CNN-BiLSTM	0.606	0.274	8.021	0.992
	TCN-BiLSTM	0.6	0.281	8.233	0.992
	PSO-SATCN-BiLSTM	0.554	0.261	7.638	0.993
	SSA-SATCN-BiLSTM	0.506	0.206	6.037	0.994
	DBO-SATCN-BiLSTM	**0.439**	**0.225**	**6.589**	**0.996**
Winter	CNN	0.644	0.347	8.689	0.991
	BiLSTM	0.626	0.332	8.291	0.992
	TCN	0.623	0.341	8.515	0.992
	CNN-BiLSTM	0.612	0.322	8.061	0.992
	TCN-BiLSTM	0.592	0.33	8.244	0.993
	PSO-SATCN-BiLSTM	0.484	0.316	6.663	0.993
	SSA-SATCN-BiLSTM	0.475	0.307	6.714	0.993
	DBO-SATCN-BiLSTM	**0.416**	**0.286**	**6.163**	**0.995**

Self-attention mechanism

Ashish Vaswani et al.[30] introduced the self-attention mechanism in 2017, and achieved a significant enhancement over the traditional attention mechanism within the encoder-decoder framework. This innovation emphasizes internal data continuity and relevance, reducing reliance on external information. Notably, it showcases a robust capacity to extract and assimilate critical data features, making it particularly apt for

Table 6. Photovoltaic farm data statistics

Solar station name	Statistics	Power output (MW)	Total solar irradiance (W/m^2)	Direct normal irradiance (W/m^2)	Global horizontal irradiance (W/m^2)	Air temperature (°C)
PV farm 1	Mean	9.7	266.4	93.3	67.7	13.1
	Minimum	0.0	0.0	0.0	0.0	−18.2
	Maximum	48.3	1359.0	980.0	989.0	41.2
	Standard deviation	13.7	368.0	200.8	111.2	14.3

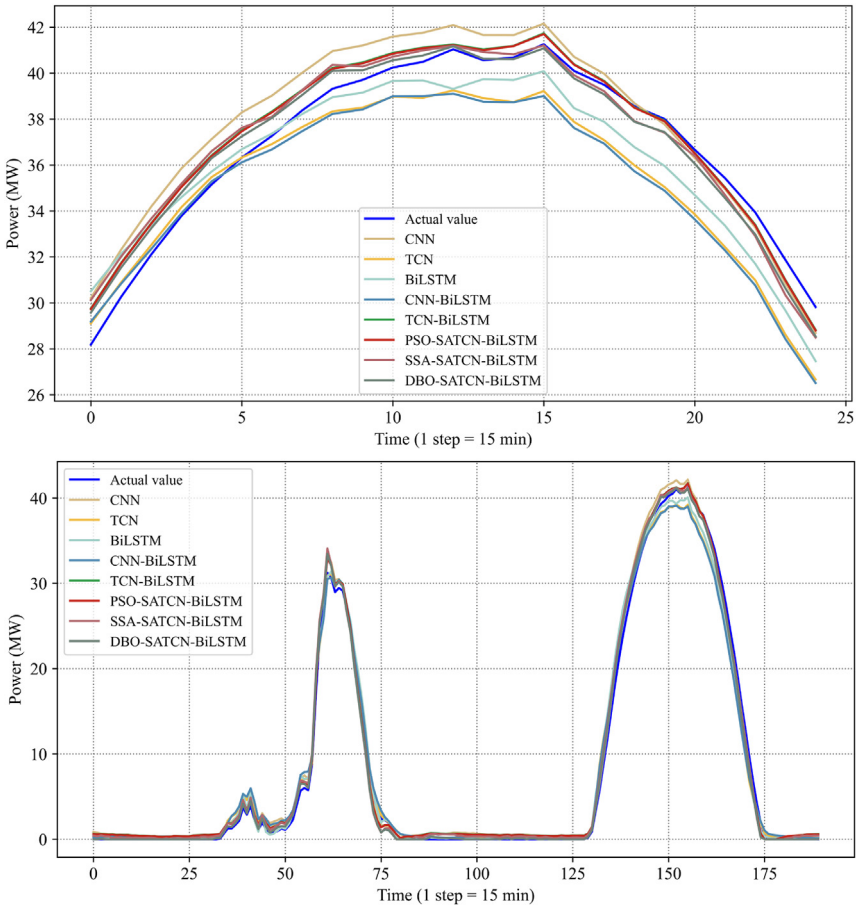


Figure 9. Public dataset single-step prediction

(A) Whole.

(B) Part.

deep learning methods handling sequential data. While LSTM and gated recurrent unit (GRU) algorithms excel at retaining long-term data, their limitation lies in overlooking internal temporal features within sequential data, lacking sensitivity to these nuances, and failing to capture inherent data connections. Their capacity to extract and assimilate crucial feature information is constrained, hindering further enhancement of model prediction accuracy. The incorporation of the self-attention mechanism serves as an effective resolution to this issue.[31]

The PV power generation process is intricately influenced by time, environmental conditions, and various other factors. This study integrates the self-attention mechanism into the prediction model, offering a robust avenue for delving into the intricate correlations within the time-series data of PV power generation. Self-attention mechanism excels at discerning patterns and correlations across different time points, enabling the model to effectively identify subtle timing discrepancies influenced by weather changes, day-night cycles, and seasonal variations. Accurate prediction of PV power necessitates a comprehensive grasp of diverse influencing factors like temperature, irradiance, cloud cover, and historical power generation records. The self-attention mechanism establishes an advanced framework to encapsulate the significance of these varied features across multiple time intervals, its dynamic allocation of distinct weights to these features enables the model to discern their interrelation and impact on current power output, and the capability facilitates more precise forecasts by recognizing and accommodating complex relationships.[32]

Figure 13 shows the architectural and directional properties of the attention mechanism, including components such as inputs, attention scores, respective distributions, and final outputs.

A deviation from the common attention mechanism, the self-attention mechanism relies solely on trainable parameters derived from the input. This approach emphasizes the inherent interdependencies within the input feature data, diminishing reliance on external information while bolstering the memory function. The formulas are as follows:

$$attention(Q, K, V) = Softmax\left(\frac{QK^T}{d_k}\right) \times V \quad \text{(Equation 2)}$$

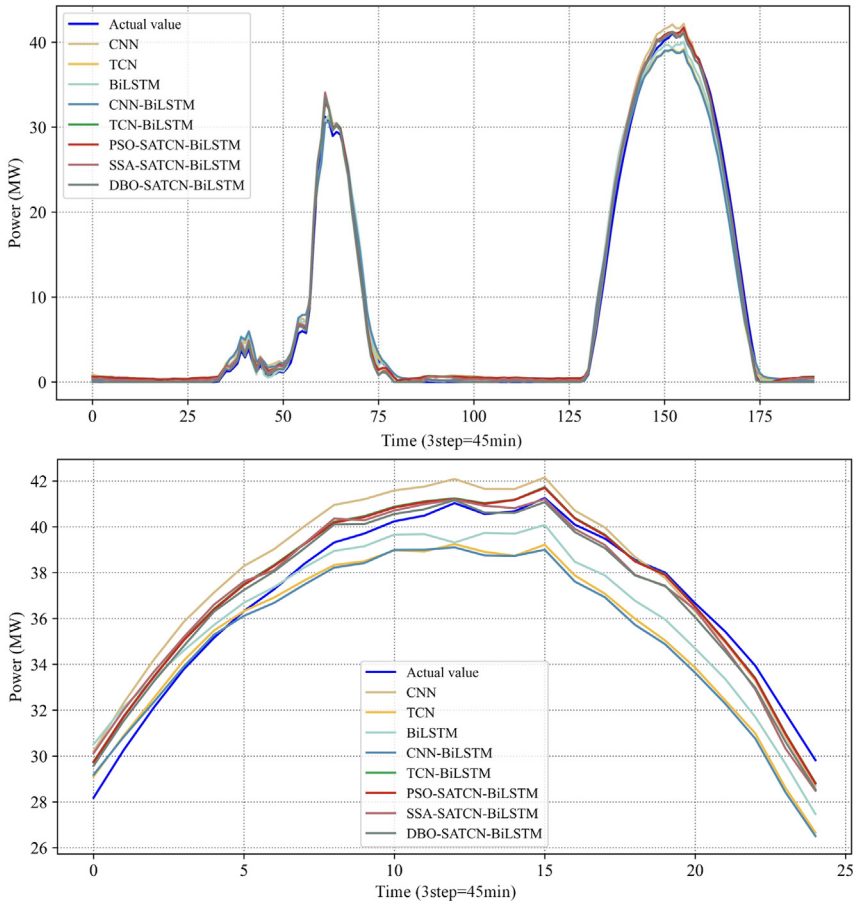


Figure 10. Public dataset multi -step prediction

where Q represents the query matrix, K stands for the key matrix, and V denotes the value matrix, obtained by different linear mappings of the same input; d_k is the scaling factor to keep the gradient stable; $softmax()$ is the normalized activation function.

Bidirectional long and short-term memory networks

Hochreiter et al.[33] introduced the LSTM network in 1997, aiming to address the issue of vanishing and exploding gradients commonly experienced in RNNs.

LSTM can handle time series problems efficiently. Illustrated in Figure 14, c_t represents the state inside the cell, and h_t denotes the output of the hidden layer. LSTMs utilize three gates—forgetting gate f_t, input gate i_t, and output gate o_t—to regulate, update, or erase information within the cell state.[34] The calculation formulas are written as follows:

$$i_t = \sigma[W_i(x_t, h_{t-1}) + b_i] \quad \text{(Equation 3)}$$

$$f_t = \sigma[W_f(x_t, h_{t-1}) + b_f] \quad \text{(Equation 4)}$$

$$o_t = \sigma[W_o(x_t, h_{t-1}) + b_o] \quad \text{(Equation 5)}$$

$$c_t = \tanh[W_c(h_{t-1}, x_t) + b_c] \quad \text{(Equation 6)}$$

$$C_t = f_t C_{t-1} + i_t C_t \quad \text{(Equation 7)}$$

$$h_t = o_t \tanh(C_t) \quad \text{(Equation 8)}$$

Table 7. Forecast error of single-step and multi-step PV annual power generation in open dataset

Step	Model	RMSE	MAE	wMAPE	R2
1 (15 min)	CNN	1.027	1.002	9.445	0.988
	BiLSTM	0.993	0.984	8.142	0.989
	TCN	0.983	0.912	7.993	0.989
	CNN-BiLSTM	0.834	0.874	7.395	0.992
	TCN-BiLSTM	0.801	0.703	7.334	0.992
	PSO-SATCN-BiLSTM	0.759	0.655	6.732	0.993
	SSA-SATCN-BiLSTM	0.752	0.651	6.695	0.994
	DBO-SATCN-BiLSTM	**0.576**	**0.438**	**5.294**	**0.994**
3 (45 min)	CNN	1.207	1.004	10.094	0.983
	BiLSTM	1.002	0.989	9.001	0.985
	TCN	1.01	0.983	8.796	0.985
	CNN-BiLSTM	0.979	0.892	8.011	0.986
	TCN-BiLSTM	0.891	0.884	7.548	0.988
	PSO-SATCN-BiLSTM	0.811	0.792	7.297	0.991
	SSA-SATCN-BiLSTM	0.805	0.775	7.344	0.99
	DBO-SATCN-BiLSTM	**0.722**	**0.613**	**6.849**	**0.991**

where W_i, W_f, W_o, and W_c represent the corresponding weight vectors; b_i, b_f, b_o, and b_c denote the bias matrices associated with the gates. h_{t-1} is the hidden layer output at time t-1, x_t is the input at time t. σ, $tanh$ is the activation function.

Training an LSTM neural network is more advantageous than training an RNN by improving the hidden layer structure and using a gating mechanism. This improvement allows for data information screening, data flow control, and retention of long-term information, effectively enhancing the ability to memorize data.

Traditional unidirectional neural networks primarily use PV data trained by front-to-back propagation techniques through time series analysis, and they have limited data utilization and fail to take full advantage of inherent data properties. In contrast, BiLSTM combines LSTM units with bi-directional processing, an architecture that enables the network to capture the dependencies of past and future data points. This is critical for tasks involving time series analysis and complex sequential data. BiLSTM is less prone to gradient vanishing problems than traditional LSTMs, allowing for more stable and efficient training.

The BiLSTM network model depicted in Figure 15 comprises two LSTM networks operating in both forward and reverse directions.[35] This bidirectional structure enables the model to forecast PV power based on data patterns before and after a given change. Compared to LSTM, BiLSTM excels in extracting intricate information from complex electric power data when predicting PV power. BiLSTM's distinct advantage is its capacity to capture features from historical and future data simultaneously while preserving the interdependency among input features. This ability makes BiLSTM more adept at extracting features from complex PV power data, allowing it to retain past and future data features while maintaining the intricate relationships between input features.

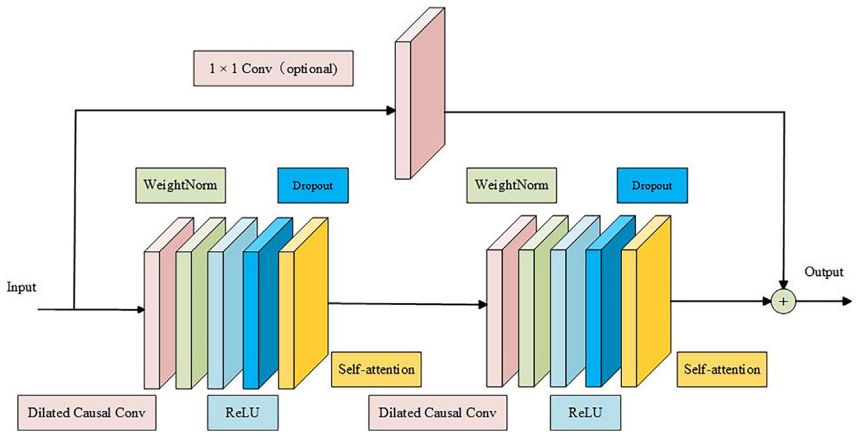


Figure 11. Schematic of SATCN network structure

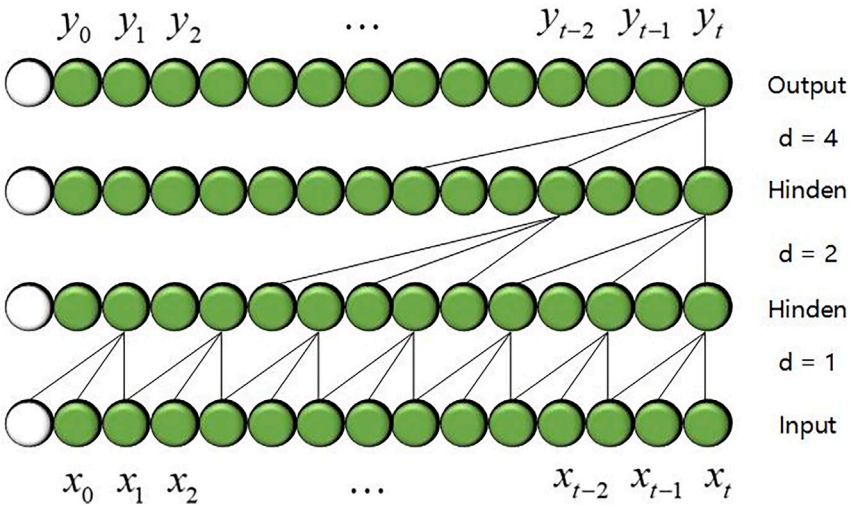


Figure 12. Dilated convolutional structure

Hyperparameter optimization using DBO algorithm

Xue et al.[36] in 2022 first introduced the DBO algorithm. The algorithm categorizes dung beetle populations into four distinct groups: ball-rolling, nest, small, and thieving dung beetles. The movements of ball-rolling dung beetles are intricately shaped by various natural conditions during their initial quest for a secure foraging site. Brood beetles strategically deposit their eggs in identified safe areas, while small dung beetles concentrate on foraging within the most optimal areas. In contrast, thieving dung beetles locate their food sources by observing the locations of other dung beetles and identifying the most advantageous feeding spots.[37] The specific process governing the update of positions for these four dung beetle categories is outlined as follows.

Rolling dung beetle. According to the phenomenon that the position and direction of the dung beetle rolling ball are affected by light, a mathematical model is constructed in which the updated position is affected by the numerical value of light intensity. The specific update method is described as follows:

$$\begin{cases} x_i(t+1) = x_i(t) + \alpha \times k \times x_i(t-1) + b \times \Delta x \\ \Delta x = |x_i(t) - X^w| \end{cases} \quad \text{(Equation 9)}$$

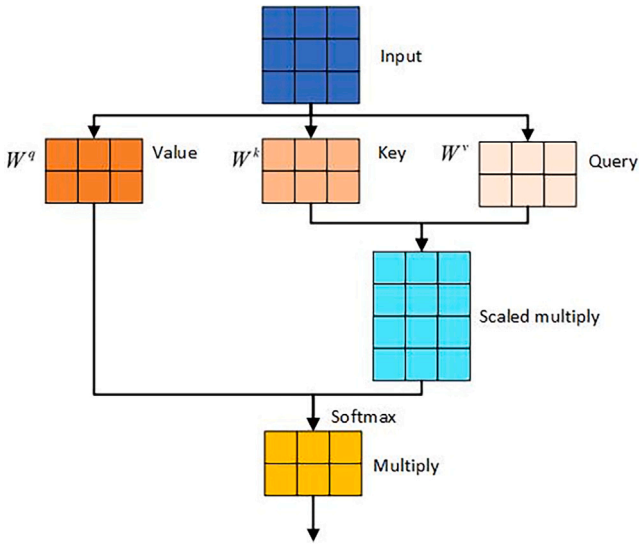


Figure 13. Computational structure of the self-attention mechanism

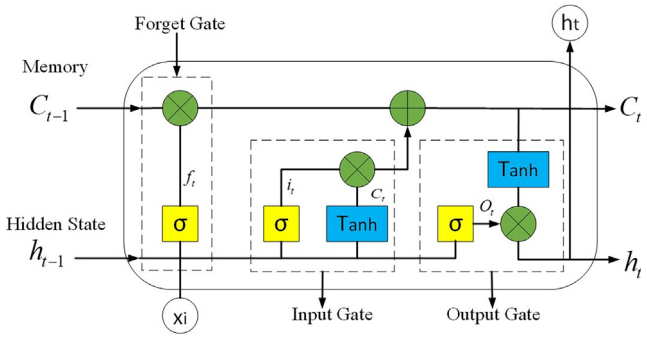


Figure 14. Structure of LSTM network neurons

where t represents the count of current iteration; a denotes the natural factor coefficient, which is determined as 1 or −1 according to the probability coefficient; the deflection coefficient k lies within the range of 0–0.2, which is usually taken as 0.1; $x_i(t)$ denotes the position information of the ith dung beetle at the tth iteration; b is a constant between 0 and 1; and X^w denotes the position in the system that is considered the global worst, while Δx is employed to simulate the alteration in light intensity.

When a dung beetle encounters an obstacle, it needs to adjust its direction by dancing to obtain a new route. The formula for updating the position is defined as follows:

$$x_i(t+1) = x_i(t) + \tan(\theta) \cdot |x_i(t) - x_i(t-1)| \quad \text{(Equation 10)}$$

where $\theta \in [0,\pi]$, the position is not updated when θ is 0, $\pi/2$ or π. It can be seen that the position update is closely related to the current position $x_i(t)$ and the historical position information $x_i(t\text{-}1)$.[38]

Breeding dung beetles. Breeding dung beetles are modeled using a boundary selection strategy to simulate female dung beetle spawning regions, which are described as follows:

$$\begin{cases} Lb^* = \max(X^* \times (1-R), Lb) \\ Ub^* = \min(X^* \times (1+R), Ub) \\ R = 1 - \dfrac{t}{T_{max}} \end{cases} \quad \text{(Equation 11)}$$

where X^* is the current local optimum position; Lb^* and Ub^* denote the lower and upper limits of the spawning area; T_{max} denotes the maximum number of iterations, R is between 0 and 1, which changes with t; and Lb and Ub denote the upper and lower limits of the optimization objective. The spawning area is constantly updated, and the position of the brood ball is changed accordingly:

$$B_i(t-1) = X^* + b_1 \times (B_i(t) - Lb^*) + b_2 \times (B_i(t) - Ub^*) \quad \text{(Equation 12)}$$

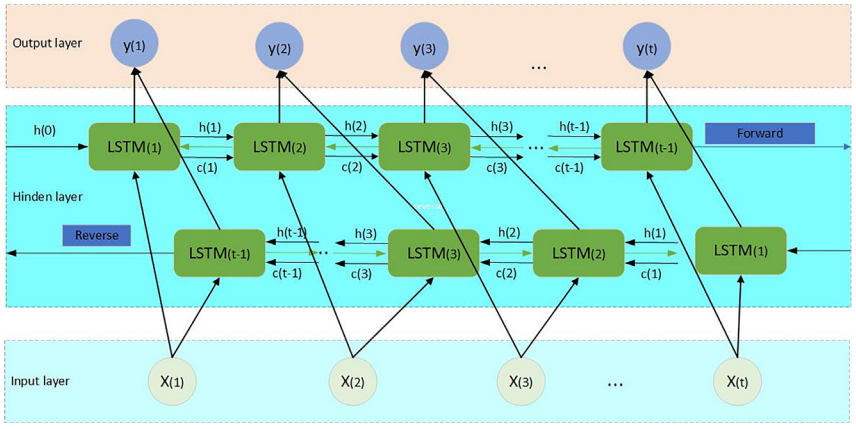


Figure 15. Schematic of BiLSTM network structure

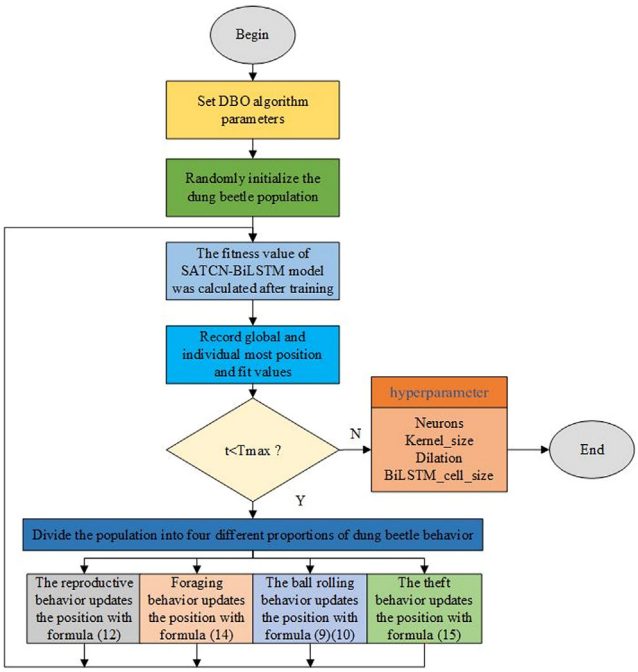


Figure 16. Flowchart of DBO optimization algorithm for SATCN-BiLSTM

where $B_i(t)$ is the positional data of the *ith* brood ball during the *tth* iteration, b_1 and b_2 represent two separate random vectors with a size of $1 \times D$, and D denotes the dimension of the optimization objective.

Toddler dung beetle. Toddler dung beetles are born and search for food in the optimal foraging area, which is expressed by the following equation:

$$Lb^b = \max\left(X^b \times (1 - R), Lb\right)$$
$$Ub^b = \min\left(X^b \times (1 + R), Ub\right) \quad \text{(Equation 13)}$$

where X^b denotes the current local optimal position, Lb^b and Ub^b denote the lower and upper limits of the optimal foraging area, and the other parameters are defined in Equation 11. The position of the toddler dung beetle is updated as follows:

$$x_i(t+1) = x_i(t) + C_1 \times \left(x_i(t) - Lb^b\right) + C_2 \times \left(x_i(t) - Ub^b\right) \quad \text{(Equation 14)}$$

where C_1 denotes a random number, $C_2 \in (0, 1)$ denotes a random vector, and the other parameters are defined in Equations 9 and 13.

Thief dung beetle. The thief dung beetle steals dung balls; X^b is the optimal point of competition for food, so the thief dung beetle's location information is updated as follows:

$$x_i(t+1) = X^b + S \times g \times \left(|x_i(t) - X^*| + |x_i(t) - X^b|\right) \quad \text{(Equation 15)}$$

where g denotes a random vector, S is a constant value, and the other parameters are defined in Equations 10 and 11.

The DBO algorithm's flow is depicted in Figure 16, outlining its primary steps as follows. (1) Setting the parameters for the DBO algorithm and population initialization. The number of dung beetle population m is determined, the problem dimension is designated as D, the count of iterations T_{max}, and the population positions are randomly initialized. (2) Calculate the fitness value of all dung beetle positions according to the objective function. (3) The population individuals are classified into ball rolling, breeding, foraging and stealing. The location of dung beetle ball rolling behavior is obtained according to Equation 10, and the local fitness value X^* after ball rolling, and the location of breeding, foraging and stealing behaviors are calculated according to Equations 12, 14, and 15. (4) Update current fitness value. (5) Repeat the aforementioned steps until the iteration is completed, and output the optimal fitness value and the global optimal solution of X^b.[39]

DBO-SATCN-BiLSTM model

The DBO-SATCN-BiLSTM model for PV power prediction is shown in Figure 17. It primarily consists of an input layer, three layers of SATCN residual blocks, two layers of BiLSTM, and a fully connected layer. The SATCN structure enhances the ability to extract temporal features from

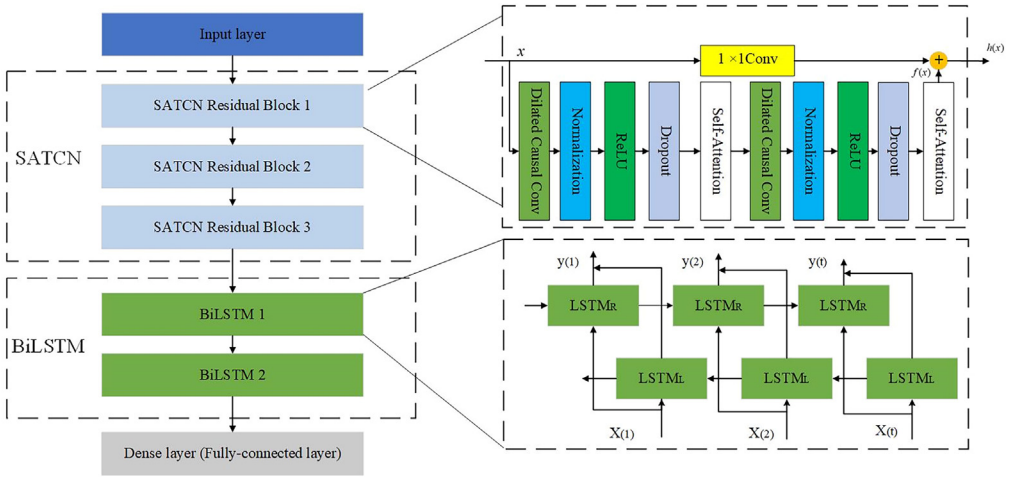


Figure 17. SATCN-BiLSTM model architecture

PV power data, BiLSTM further captures the temporal correlation between forward and backward features, and the fully connected layer facilitates dimensional conversion. Figure 18 illustrates the overall experimental process of the model, the information changes between models, and the evaluation indicators used in this paper. The overall model consists of SATCN and BiLSTM, with DBO providing hyperparameter optimization. The complete workflow of the model is as follows.

PV-related data are obtained directly from the power plant; the original dataset includes several features with weak correlation to PV output power. To accurately identify the features necessary for the model, a correlation analysis will be conducted prior to model training. Environmental factors with higher correlation to PV power output will be selected as feature data for prediction. Subsequently, the selected feature factors will undergo further cleaning to remove outliers and erroneous values.

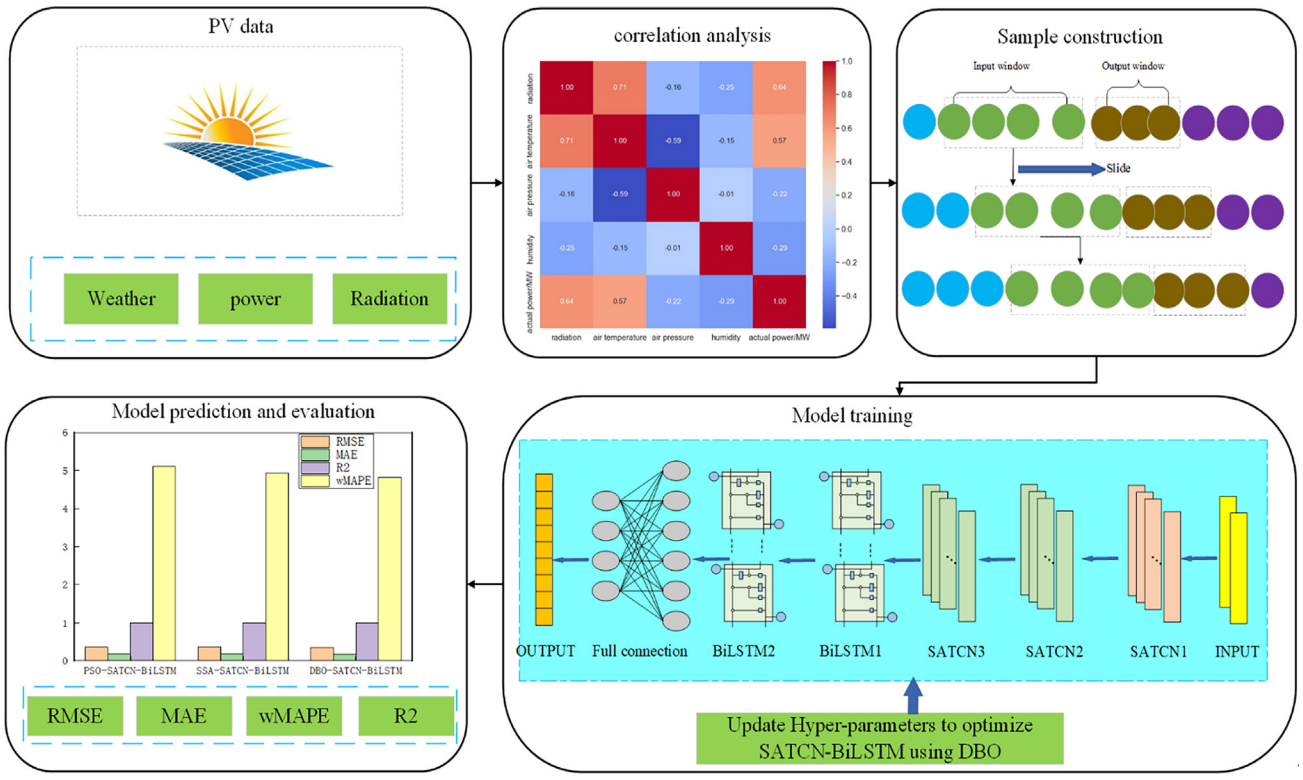


Figure 18. DBO-SATCN-BiLSTM workflow diagram

Before the model makes a real prediction for a certain area, the historical data of this area is fed into the hybrid model containing the DBO optimization algorithm, which simulates the behaviors of dung beetles such as foraging, reproduction, rolling, and stealing, and optimizes the hyperparameters of the SATCN-BiLSTM hybrid model. Hyperparameter optimization continuously iterates the hybrid model in historical data, records the evaluation indicators after each run, and reversely adjusts the hyperparameter values in the hybrid model according to the trend of the evaluation indicators. The behavior of adjusting hyperparameters is essentially to adjust the calculation method of updating feature weights of the hybrid model. Compared with other existing optimization algorithms, such as SSA and PSO, the DBO algorithm has 4 optimization methods. More optimization methods mean that the DBO algorithm can more comprehensively perceive the value of the best hyperparameters. After determining the model hyperparameters, when predicting the data of the same power plant, the model keeps the hyperparameters unchanged.

When the hybrid model makes a real prediction, the time features contained in the data are first extracted through the three-layer SATCN structure. SATCN combines the self-attention mechanism and the temporal convolutional network; self-attention is used to capture the dependencies between different time steps in time series data, while temporal convolution is used to learn local and global temporal features. Compared with the traditional TCN network, SATCN can expand the receptive field without increasing the number of parameters through the dilated convolution structure, effectively capturing the long-term dependencies in the input sequence. By combining self-attention with TCN, the model can effectively integrate and utilize multimodal data sources, process long sequences, and learn to weigh the impact of weather conditions in different ways according to the time of day or season, thereby improving the prediction accuracy in different environmental contexts.

After the SATCN layer, the two BiLSTM layers will use their bidirectional data processing capabilities to obtain the forward and backward data change characteristics in the sequence data. The forward LSTM is responsible for capturing the information before the time step, while the backward LSTM captures the information after the time step. The two use the bidirectional information through the back propagation and merging of hidden states. Due to the existence of the bidirectional structure, BiLSTM has higher stability than LSTM when facing drastic data changes. The model will convert the high-dimensional data into one-dimensional prediction data through a fully connected layer, and then obtain the predicted power value through subsequent operations such as inverse normalization.

As for the proposed DBO-SATCN-BiLSTM hybrid model, DBO solves the problem of unbalanced model weight updates and enhances the model's ability to calculate the intrinsic relationship between features and power, SATCN effectively captures the temporal characteristics of the sequence through self-attention and temporal convolution, and BiLSTM pays more attention to the changing trend of data in the bidirectional dimension and retains useful information in the long-term dimension. The combination can improve the quality of model parameters and make the advantages and avoid disadvantages of each model complementary, thereby maximizing the accuracy and stability in the face of complex weather conditions.

Data preprocessing and model evaluation

Data preprocessing. The dataset contains information gathered a year from a 30 MW PV power plant located in Shaanxi Province, China. Figure 19 showcases the data types encompassing power, solar radiation, and weather information. The specific geographical location of the plant is depicted in Figure 20, with coordinates at a longitude of 110°30′ and a latitude of 38°49′. Each measurement sample within the dataset follows a 15-min time interval, covering the duration from January 1st, 2021, to December 31st, 2021.

In the original dataset, every parameter or variable exhibits a unique scale. To mitigate magnitude discrepancies and expedite the training process, normalization techniques are applied to the wind and light power data, converting them into scalar values.[40] Specifically, transform the data into the range [0, 1], following the equation as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad \text{(Equation 16)}$$

where x' represents the normalized value; x denotes the actual value of the sample; x_{max} and x_{min} stand for the max-min values within the columns where the data are located.

PV power generation is influenced by multiple factors, with weather factors, mainly irradiance, being the most significant among them. The weather data are used as input to improve forecast accuracy. The weather factors in the raw data contain properties such as temperature, humidity, barometric pressure, and solar radiation. Not all of these weather factors have an equally significant impact on PV power generation, which can lead to redundancy in the input features. This redundancy not only increases the time required for model training, but may also reduce the accuracy of the predictions. The input features are correlated by using the Pearson correlation coefficient for screening and optimization.

Pearson's correlation coefficient is a statistical measure utilized to assess the extent of linear correlation between two random variables. Also recognized as product-moment correlation, it assesses both the intensity and direction of a linear relationship. When one variable consistently increases or decreases in tandem with another, the covariance of the two variables is positive. Conversely, if they move in opposite directions, the covariance is negative.[41] The formula for covariance is calculated by the following equation:

$$\rho(x,y) = \frac{\sum_{i=1}^{n}\left((x_i - x_\mu)\left(y_i - y_\mu\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - x_\mu\right)^2 \sum_{i=1}^{n}\left(y_i - y_\mu\right)^2}} \quad \text{(Equation 17)}$$

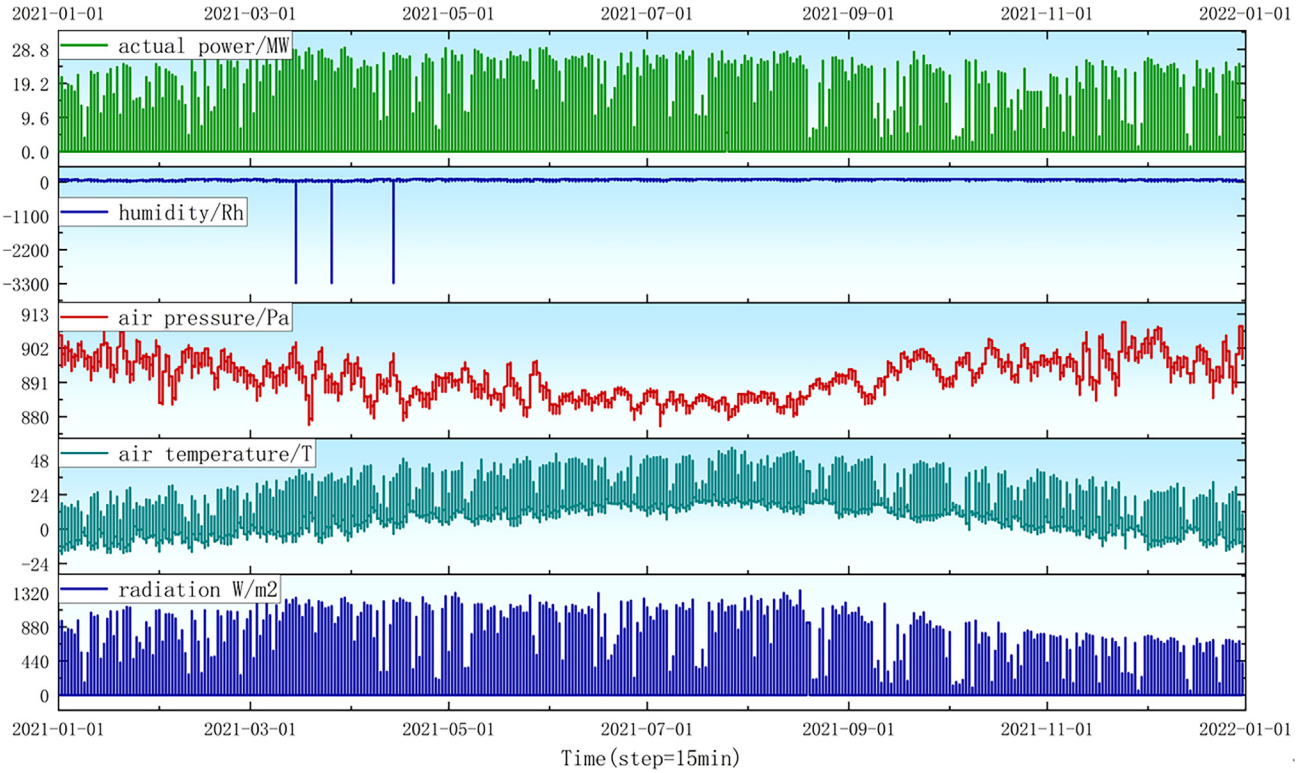


Figure 19. Raw data of the PV power plant

where $\rho(x,y)$ represents Pearson correlation coefficient between x and y, ranging from −1 to 1; x_μ and y_μ represent the average values of x and y, and when the resulting value approaches 1, it indicates a strong positive correlation between the two sets of variables; when the obtained value is close to −1, it means that there is a high degree of negative correlation; when the obtained value is close to 0, there is a low degree of linear correlation. The correlation coefficients for raw PV power data are shown in Figure 21, revealing that temperature and radiance have the strongest correlation, while humidity and pressure show negative correlation.

Figure 20. Location of the PV power plant

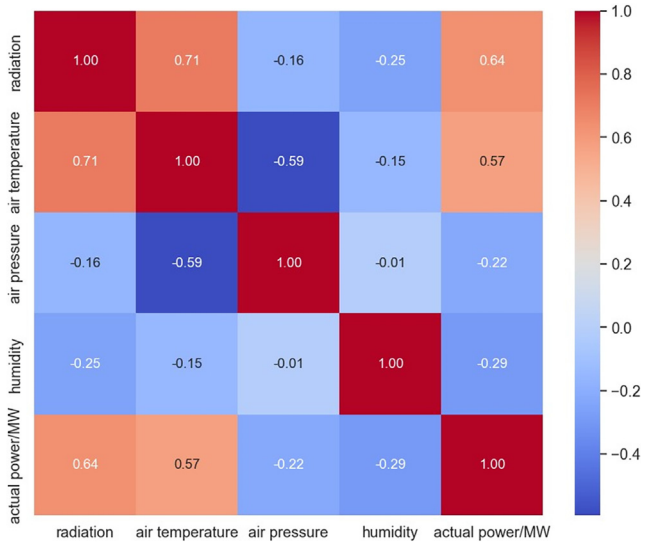


Figure 21. Matrix of characteristic correlation coefficients for raw PV power data

In this research, the time series-based neural network was used for PV power prediction and the sliding window algorithm to serialize the power data operation. On the one hand, it can divide the data into input features and target features, converting them into supervised learning. On the other hand, it can transform the input data into three dimensions, realizing dynamic prediction.[42] Figure 22 shows the data before and after going through the sliding window, the input sequence consists of 4 elements, and the output comprises 3 elements. The dataset is split into a training set and a test set following an 8:2 ratio.

Model evaluation indicators. To validate the model's efficacy, four distinct assessment indices are employed: RMSE, MAE, wMAPE, and R2. These metrics measure the disparity between predicted and actual values, providing a measure of the model's predictive performance. This evaluation is formulated as follows:[43]

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y(i)_{pred} - y(i)_{test}\right)^2} \quad \text{(Equation 18)}$$

$$wMAPE = \frac{1}{n}\frac{\sum_{i=1}^{n}\left|y(i)_{pred} - y(i)_{test}\right|}{\sum_{i=1}^{n} y(i)_{test}} \quad \text{(Equation 19)}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y(i)_{pred} - y(i)_{test}\right| \quad \text{(Equation 20)}$$

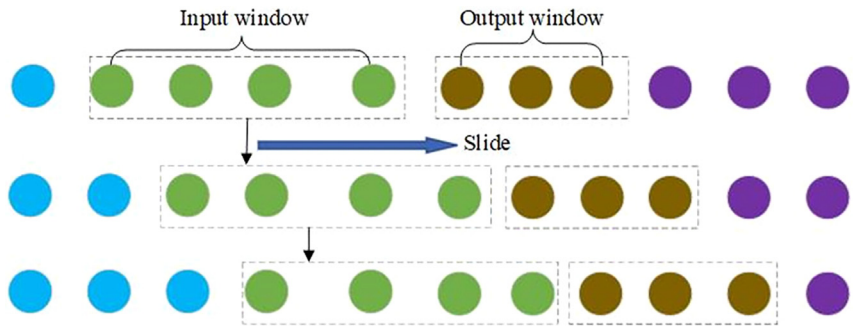


Figure 22. Sliding window schematic

Nomenclature

Acronyms

CNN Convolutional Neural Networks
BP Back Propagation
SVM Support Vector Machines
TCN Temporal Convolutional Network
SATCN Self-Attention-TCN
GRU Gated Recurrent Unit
LSTM Long Short-Term Memory
BiLSTM Bi-directional Long Short-term Memory
RNN Artificial Neural Networks
GA Genetic Algorithm
PSO Particle Swarm Optimization
DBO Dung Beetle Optimization
SSA Sparrow Search Algorithm
CSO Cat Swarm Optimization
ACO Ant Colony Optimization
MAE Mean Absolute Error
RMSE Root Mean Square Error
wMAPE Weighted Mean Absolute Percentage Error
R2 Coefficient of Determination

Variables

$F(s)$ Extended convolution
k Convolution kernel size
$f(i)$ The ith element
$f(t)$ Forget gate
C_t Cell state
i_t Input gate
o_t Output gate
h_t Hidden layer output
Q Query matrix
K Key matrix
V Value matrix
σ,tanh Sigmoid activation function
$x_i(t)$ Dung beetle location information
Lb^b, Ub^b Lower and Upper boundaries of search space
T_{max} Maximum iterations
X^b Local optimum position
x' Data normalization
x_{max}, x_{min} Data maximum and minimum values
$\rho(x,y)$ Pearson correlation coefficient
n Number of predicted samples
y_{test} Actual power value
y_{pred} Predicted power value

$$R2 = 1 - \frac{\sum_{i=1}^{n}\left(y(i)_{pred} - y(i)_{test}\right)^2}{\sum_{i=1}^{n}\left(y(i)_{mtest} - y(i)_{test}\right)^2} \quad \text{(Equation 21)}$$

where n represents the count of predicted samples; y_{test} is the actual PV power value; y_{pred} is the predicted PV power value; and y_{mtest} is the average of the actual value of PV power. RMSE, wMAPE, and MAE all measure how close the predicted value is to the actual value, with smaller values indicating a more accurate prediction. At the same time, R2 represents the fraction of the dependent variable elucidated by the independent variable. As R2 approaches 1, it denotes a higher degree of alignment between the model and the data, signifying that the predicted values closely match the actual observations.

Conclusion

In this study, the DBO-SATCN-BiLSTM model was introduced to further enhance PV power prediction accuracy by incorporating weather factors. This model utilizes SATCN to extract temporal features from the data and integrates BiLSTM to establish correlations between features

and power output. Comparative analyses were conducted against several models including CNN, TCN, BiLSTM, CNN-BiLSTM, TCN-BiLSTM hybrid models, PSO-SATCN-BiLSTM, and SSA-SATCN-BiLSTM. Results indicate that the proposed model undergoes evaluation for both single-step and multi-step prediction across various seasons, showcasing superior prediction accuracy and an ability to capture intricate temporal patterns in PV power generation data compared to the other seven models.

In single-step prediction, the DBO-SATCN-BiLSTM achieves an RMSE of 0.357 when predicting one year's PV power. This demonstrates a significant reduction of 33.1%, 23.4%, and 18.1% compared to the CNN, TCN, and BiLSTM. It outperforms CNN-BiLSTM, TCN-BiLSTM, PSO-SATCN-BiLSTM, and SSA-SATCN-BiLSTM by margins of 17.5%, 10%, 2.9%, and 2.4%. Both MAE and wMAPE of the proposed model consistently remain lower across different seasons compared to the baseline models.

In terms of multi-step (3-step) prediction, as the prediction range increases, errors escalate across all models. The DBO-SATCN-BiLSTM stands out with an RMSE of 0.437, indicating substantial improvements by 52.3%, 32.4%, 32.9%, 31.5%, 31.1%, 9.5%, and 4.7% when compared with the other seven models. Besides, its MAE, wMAPE, and R2 metrics is 0.217, 5.966, and 0.995, surpassing those of the CNN, TCN, BiLSTM, CNN-BiLSTM, TCN-BiLSTM, PSO-SATCN-BiLSTM, and SSA-SATCN-BiLSTM. Across different seasons, the DBO-SATCN-BiLSTM consistently outperforms the contrast models inform the aspect of RMSE, MAE, wMAPE, and R2 in multi-step prediction scenarios.

In conclusion, experimental results strongly support the high accuracy of the proposed DBO-SATCN-BiLSTM model compared to other different models in both single-step and multi-step prediction contexts. The suggested approach offers an understanding for accurately forecasting PV power systems. Despite extensive efforts, limitations and challenges persist in this study. The proposed prediction model incorporates several different network layers, increasing the network's complexity. Although a dropout layer has been added, computation time remains lengthy. The next phase of research will aim to carefully scrutinize and overcome these obstacles and constraints.

Limitations of the study

Despite extensive efforts, limitations and challenges persist in this study are as follows: the prediction model proposed in this paper is connected by several different network layers, which increases the complexity of the network to some extent, and although the dropout layer is added, there is still a long computation time. The following phase of research will strive to carefully scrutinize and overcome these obstacles and constraints.

RESOURCE AVAILABILITY

Lead contact

Further requests for information should be directed and will be handled by the corresponding author and lead contact, Zhizhuo Qiu (102210382@hbut.edu.cn).

Materials availability

The study did not generate new materials.

Data and code availability

- The input data are stored on GitHub, including the raw data and the code that processes the raw data, these accession numbers for the datasets are listed in the key resources table.
- All original code has been deposited at Github and is publicly available as of the date of publication. DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This paper was supported by the National Natural Science Foundation of China (51977061 and 51407063) and the Open Foundation of Hubei Key Laboratory for High-efficiency Utilization of Solar Energy and Operation Control of Energy Storage System (HBSEES202205).

AUTHOR CONTRIBUTIONS

Formal analysis and funding acquisition, R.Q.; conceptualization and methodology, Z.Q. and R.Q.; writing – original draft and visualization, Z.Q. and H.W.; writing – review & editing and validation, R.Q., H.W., Z.Y., and X.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Pre-processing of data
  - Model parameter optimization
  - SATCN-BiLSTM model for photovoltaic power prediction

- Dropout layer
- QUANTIFICATION AND STATISTICAL ANALYSIS

Received: April 14, 2024
Revised: August 10, 2024
Accepted: October 4, 2024
Published: October 11, 2024

REFERENCES

1. Allouhi, A., Rehman, S., Buker, M.S., and Said, Z. (2022). Up-to-date literature review on solar PV systems: Technology progress, market status and R&D. J. Clean. Prod. *362*, 132339. https://doi.org/10.1016/j.jclepro.2022.132339.
2. Lin, X.M., Kireeva, N., Timoshin, A.V., Naderipour, A., Abdul-Malek, Z., and Kamyab, H. (2021). A multi-criteria framework for designing of stand-alone and grid-connected photovoltaic, wind, battery clean energy system considering reliability and economic assessment. Energy *224*, 120154. https://doi.org/10.1016/j.energy.2021.120154.
3. Nasri, S., Zamanifar, M., Naderipour, A., Nowdeh, S.A., Kamyab, H., and Abdul-Malek, Z. (2023). Stability and dynamic analysis of a grid-connected environmentally friendly photovoltaic energy system. Environ. Sci. Pollut. Res. Int. *30*, 71701–71713. https://doi.org/10.1007/s11356-021-15255-w.
4. Mayer, M.J., and Yang, D. (2022). Probabilistic photovoltaic power forecasting using a calibrated ensemble of model chains. Renew. Sustain. Energy Rev. *168*, 112821. https://doi.org/10.1016/j.rser.2022.112821.
5. Naderipour, A., Nowdeh, S.A., Babanezhad, M., Najmi, E.S., Kamyab, H., and Abdul-Malek, Z. (2023). Technical-economic framework for designing of water pumping system based on photovoltaic clean energy with water storage for drinking application. Environ. Sci. Pollut. Res. Int. *30*, 71754–71765. https://doi.org/10.1007/s11356-021-16342-8.
6. Markovics, D., and Mayer, M.J. (2022). Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew. Sustain. Energy Rev. *161*, 112364. https://doi.org/10.1016/j.rser.2022.112364.
7. Mayer, M.J. (2022). Benefits of physical and machine learning hybridization for photovoltaic power forecasting. Renew. Sustain. Energy Rev. *168*, 112772. https://doi.org/10.1016/j.rser.2022.112772.
8. Fjelkestam Frederiksen, C.A., and Cai, Z. (2022). Novel machine learning approach for solar photovoltaic energy output forecast using extra-terrestrial solar irradiance. Appl. Energy *306*, 118152. https://doi.org/10.1016/j.apenergy.2021.118152.
9. Haider, S.A., Sajid, M., Sajid, H., Uddin, E., and Ayaz, Y. (2022). Deep learning and statistical methods for short-and long-term solar irradiance forecasting for Islamabad. Renew. Energy *198*, 51–60. https://doi.org/10.1016/j.renene.2022.07.136.
10. Nguyen, T.N., and Müsgens, F. (2022). What drives the accuracy of PV output forecasts? Appl. Energy *323*, 119603. https://doi.org/10.1016/j.apenergy.2022.119603.
11. Ma, M., He, B., Shen, R., Wang, Y., and Wang, N. (2022). An adaptive interval power forecasting method for photovoltaic plant and its optimization. Sustain. Energy Technol. Assessments *52*, 102360. https://doi.org/10.1016/j.seta.2022.102360.
12. Li, J., and Liu, Q. (2022). Forecasting of short-term photovoltaic power generation using combined interval type-2 Takagi-Sugeno-Kang fuzzy systems. Int. J. Electr. Power Energy Syst. *140*, 108002. https://doi.org/10.1016/j.ijepes.2022.108002.
13. Liang, L., Su, T., Gao, Y., Qin, F., and Pan, M. (2023). FCDT-IWBOA-LSSVR: An innovative hybrid machine learning approach for efficient prediction of short-to-mid-term photovoltaic generation. J. Clean. Prod. *385*, 135716. https://doi.org/10.1016/j.jclepro.2022.135716.
14. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R.M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imag. *35*, 1285–1298. https://doi.org/10.1109/TMI.2016.2528162.
15. Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys. Nonlinear Phenom. *404*, 132306. https://doi.org/10.1016/j.physd.2019.132306.
16. Ewees, A.A., Al-qaness, M.A., Abualigah, L., and Elaziz, M.A. (2022). HBO-LSTM: Optimized long short term memory with heap-based optimizer for wind power forecasting. Energy Convers. Manag. *268*, 116022. https://doi.org/10.1016/j.enconman.2022.116022.
17. Bai, S., Kolter, J.Z., and Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. Preprint at arXiv. https://doi.org/10.48550/arXiv.1803.01271.
18. Quan, R., Zhang, J., Li, X., Guo, H., Chang, Y., and Wan, H. (2024). A hybrid CNN–BiLSTM–AT model optimized with enhanced whale optimization algorithm for remaining useful life forecasting of fuel cell. AIP Adv. *14*, 025251. https://doi.org/10.1063/5.0191483.
19. Khan, S., Fazil, M., Sejwal, V.K., Alshara, M.A., Alotaibi, R.M., Kamal, A., and Baig, A.R. (2022). BiCHAT: BiLSTM with deep CNN and hierarchical attention for hate speech detection. J. Inflamm. Res. *34*, 4335–4344. https://doi.org/10.1016/j.jksuci.2022.05.006.
20. Zou, M., Holjevac, N., Đaković, J., Kuzle, I., Langella, R., Giorgio, V.D., and Djokic, S.Z. (2022). Bayesian CNN-BiLSTM and vine-GMCM based probabilistic forecasting of hour-ahead wind farm power outputs. IEEE Trans. Sustain. Energy *13*, 1169–1187. https://doi.org/10.1109/TSTE.2022.3148718.
21. Limouni, T., Yaagoubi, R., Bouziane, K., Guissi, K., and Baali, E.H. (2023). Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model. Renew. Energy *205*, 1010–1024. https://doi.org/10.1016/j.renene.2023.01.118.
22. Quan, R., Liang, W., Wang, J., Li, X., and Chang, Y. (2024). An enhanced fault diagnosis method for fuel cell system using a kernel extreme learning machine optimized with improved sparrow search algorithm. Int. J. Hydrogen Energy *50*, 1184–1196. https://doi.org/10.1016/j.ijhydene.2023.10.019.
23. Liu, S., Wang, L., Jiang, H., Liu, Y., and You, H. (2022). Wind farm energy storage system based on cat swarm optimization–backpropagation neural network wind power prediction. Front. Energy Res. *10*, 850295. https://doi.org/10.3389/fenrg.2022.850295.
24. Wu, Z., Cui, N., Gong, D., Zhu, F., Li, Y., Xing, L., Wang, Z., Zhu, B., Chen, X., Wen, S., and Zha, Y. (2023). Predicting daily global solar radiation in various climatic regions of China based on hybrid support vector machines with meta-heuristic algorithms. J. Clean. Prod. *385*, 135589. https://doi.org/10.1016/j.jclepro.2022.135589.
25. Abou Houran, M., Salman Bukhari, S.M., Zafar, M.H., Mansoor, M., and Chen, W. (2023). COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications. Appl. Energy *349*, 121638. https://doi.org/10.1016/j.apenergy.2023.121638.
26. Chen, Y., and Xu, J. (2022). Solar and wind power data from the Chinese state grid renewable energy generation forecasting competition. Sci. Data *9*, 577. https://doi.org/10.1038/s41597-022-01696-6.
27. Sabadus, A., Blaga, R., Hategan, S.M., Calinoiu, D., Paulescu, E., Mares, O., Boata, R., Stefu, N., Paulescu, M., and Badescu, V. (2024). A cross-sectional survey of deterministic PV power forecasting: Progress and limitations in current approaches. Renew. Energy *226*, 120385. https://doi.org/10.1016/j.renene.2024.120385.
28. Ahmed, R., Sreeram, V., Mishra, Y., and Arif, M. (2020). A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renew. Sustain. Energy Rev. *124*, 109792. https://doi.org/10.1016/j.rser.2020.109792.
29. Xiang, L., Liu, J., Yang, X., Hu, A., and Su, H. (2022). Ultra-short term wind power prediction applying a novel model named SATCN-LSTM. Energy Convers. Manag. *252*, 115036. https://doi.org/10.1016/j.enconman.2021.115036.
30. Shaw, P., Uszkoreit, J., and Vaswani, A. (2018). Self-attention with relative position representations. Preprint at arXiv. https://aclanthology.org/N18-2074.
31. Wei, Y., Wu, D., and Terpenny, J. (2023). Bearing remaining useful life prediction using

self-adaptive graph convolutional networks with self-attention mechanism. Mech. Syst. Signal Process. *188*, 110010. https://doi.org/10.1016/j.ymssp.2022.110010.

32. Wu, Y., and Zhou, J.T. (2023). A neighborhood-aware graph self-attention mechanism-based pre-training model for Knowledge Graph Reasoning. Inf. Sci. *647*, 119473. https://doi.org/10.1016/j.ins.2023.119473.
33. Wan, A., Chang, Q., AL-Bukhaiti, K., and He, J. (2023). Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism. Energy *282*, 128274. https://doi.org/10.1016/j.energy.2023.128274.
34. Quan, R., Liu, P., Li, Z., Li, Y., Chang, Y., and Yan, H. (2023). A multi-dimensional residual shrinking network combined with a long short-term memory network for state of charge estimation of Li-ion batteries. J. Energy Storage *57*, 106263. https://doi.org/10.1016/j.est.2022.106263.
35. Zhen, H., Niu, D., Wang, K., Shi, Y., Ji, Z., and Xu, X. (2021). Photovoltaic power forecasting based on GA improved Bi-LSTM in microgrid without meteorological information. Energy *231*, 120908. https://doi.org/10.1016/j.energy.2021.120908.
36. Xue, J., and Shen, B. (2023). Dung beetle optimizer: A new meta-heuristic algorithm for global optimization. J. Supercomput. *79*, 7305–7336. https://doi.org/10.1007/s11227-022-04959-6.
37. Li, Y., Sun, K., Yao, Q., and Wang, L. (2024). A dual-optimization wind speed forecasting model based on deep learning and improved dung beetle optimization algorithm. Energy *286*, 129604. https://doi.org/10.1016/j.energy.2023.129604.
38. Ratnakumar, R., and Nanda, S.J. (2021). A high speed roller dung beetles clustering algorithm and its architecture for real-time image segmentation. Appl. Intell. *51*, 4682–4713. https://doi.org/10.1007/s10489-020-02067-7.
39. Zhu, F., Li, G., Tang, H., Li, Y., Lv, X., and Wang, X. (2024). Dung beetle optimization algorithm based on quantum computing and multi-strategy fusion for solving engineering problems. Expert Syst. Appl. *236*, 121219. https://doi.org/10.1016/j.eswa.2023.121219.
40. Hosseini, P., Taheri, S., Akhavan, J., and Razban, A. (2023). Privacy-preserving federated learning: Application to behind-the-meter solar photovoltaic generation forecasting. Energy Convers. Manag. *283*, 116900. https://doi.org/10.1016/j.enconman.2023.116900.
41. Liu, X., Liu, Y., Kong, X., Ma, L., Besheer, A.H., and Lee, K.Y. (2023). Deep neural network for forecasting of photovoltaic power based on wavelet packet decomposition with similar day analysis. Energy *271*, 126963. https://doi.org/10.1016/j.energy.2023.126963.
42. Yin, L., Cao, X., and Liu, D. (2023). Weighted fully-connected regression networks for one-day-ahead hourly photovoltaic power forecasting. Appl. Energy *332*, 120527. https://doi.org/10.1016/j.apenergy.2022.120527.
43. Feroz Mirza, A., Mansoor, M., Usman, M., and Ling, Q. (2023). Hybrid Inception-embedded deep neural network ResNet for short and medium-term PV-Wind forecasting. Energy Convers. Manag. *294*, 117574. https://doi.org/10.1016/j.enconman.2023.117574.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
PV power	Chen	https://doi.org/10.1038/s41597-022-01696-6
PV data	Chen input data	https://github.com/Bob05757/Renewable-energy-generation-input-feature-variables-analysis.
Code and data for development and evaluation	This paper	https://doi.org/10.7910/DVN/ONVZPH
LSTM	Hochreiter, S et al.	https://doi.org/10.1162/neco.1997.9.8.1735
TCN	Shaojie B et al.	https://doi.org/10.48550/arXiv.1907.00235
CNN	Hoo-Chang Shin et al.	https://doi.org/10.1109/TMI.2016.2528162
Software and algorithms		
TensorFlow 2.3	TensorFlow	tensorflow (RRID:SCR_016345)
Python 3.8	Python	Python Programming Language (RRID:SCR_008394)

METHOD DETAILS

The Python programming language to implement the methods described in model building, validation, and predictive modeling. The model is built using several open source Python packages including tensorflow and Scikit-Learn, both of which are referenced in the key resources table. Complete code input data are available online.

Pre-processing of data

In this study, linear interpolation was used to interpolate the missing data to reduce the influence of discontinuous data on the temporal correlation. On the other hand, different parameters have different scales, and larger values mask the effects of smaller inputs. Therefore, the data are preprocessed using min-max normalization or standardization. The calculation is as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad \text{(Equation 22)}$$

Model parameter optimization

In this experiment, when optimizing the SARCN-BiLSTM model and other comparison models, the objective function is defined as the mean square error, and the convergence index is the maximum number of iterations. In the model experiment in which DBO participated, the population proportions of the rolling dung beetle, the brooding dung beetle, the small dung beetle, and the thief dung beetle were set to 0.4, 0.6, 0.4, and 0.4, respectively. The number of layers of SATCN was set to 3, the number of filters in each layer was set to 83, 119, and 96, the number of kernels was set to 36, 12, and 34, and the expansion index was set to 6, 3, and 2, respectively. The number of layers of BiLSTM was set to 2, and the number of BiLSTM cells in each layer was set to 23 and 199, respectively. Except for the number of layers, all hyperparameters (including the number of filters, the number of kernels, and the number of BiLSTM cells) were set by the optimization algorithm.

SATCN-BiLSTM model for photovoltaic power prediction

In the SASpaTCN-BiLSTM model, the self-attention mechanism is used to improve the TCN layer, and each TCN layer is added with the self-attention mechanism. In the attention mechanism, let a certain sequence data be $X = [x_1, x_2, x_3, \ldots, x_n]$, and the processing steps are as follows: first, the Query, Key, and Value of each element are calculated, as defined in Equations 23, 24, and 25, where W^Q, W^K, and W^V are the linear transformation weight matrices of Query, Key, and Value, respectively, and Q, K, and V represent the Query, Key, and Value matrices, respectively.

$$Q = W^Q * x_i \quad \text{(Equation 23)}$$

$$K = W^K * x_i \quad \text{(Equation 24)}$$

$$V = W^V * x_i \quad \text{(Equation 25)}$$

In the proposed model, SATCN is located at the front end of the BiLSTM layer, and the data flows from SATCN to BiLSTM. The fully connected layer finally obtains the final prediction result by processing the high-dimensional output of the BiLSTM layer.

Dropout layer

The dropout layer in a neural network serves as a form of regularization, helping to prevent overfitting. Overfitting occurs when a model learns to memorize the training data rather than generalize from it. Dropout provides a simple way to prevent this by randomly setting a fraction of input units to zero during each update of the training phase.

QUANTIFICATION AND STATISTICAL ANALYSIS

The experiment of this study is based on Pycharm2020 and the drawing tool is origin2016.The statistical details of all experiments can be found in Table 6, including the mean, maximum, minimum, and standard deviation of various statistics in the used datasets.The statistical tests used in this paper include root-mean-square error (RMSE), MAE, wMAPE and coefficient of determination (R2). The calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y(i)_{pred} - y(i)_{test}\right)^2} \quad \text{(Equation 26)}$$

$$wMAPE = \frac{1}{n}\frac{\sum_{i=1}^{n}\left|y(i)_{pred} - y(i)_{test}\right|}{\sum_{i=1}^{n} y(i)_{test}} \quad \text{(Equation 27)}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y(i)_{pred} - y(i)_{test}\right| \quad \text{(Equation 28)}$$

$$R2 = 1 - \frac{\sum_{i=1}^{n}\left(y(i)_{pred} - y(i)_{test}\right)^2}{\sum_{i=1}^{n}\left(y(i)_{mtest} - y(i)_{test}\right)^2} \quad \text{(Equation 29)}$$

Among them, *n* represents the total number of samples involved in the prediction, and its value is the number of test set samples, accounting for 20% of the total number of dataset samples, that is, 14035, *ytest* is the actual PV power value; *ypred* is the predicted PV power value; and *ymtest* is the average of the actual value of PV power.